22 pr 1>

ENHANCED PHAGE DISPLAY LIBRARIES OF HUMAN VH FRAGMENTS AND METHODS FOR PRODUCING SAME

Field of the Invention

The present invention relates to combinatorial libraries including phage display libraries which display binding fragments having preferred characteristics of solubility. The present invention also describes methods of producing phage libraries in which the phage population displays binding fragments having characteristics which are biased towards characteristics of the wild type or parental binding fragment.

Background of the Invention

Developments in antibody engineering and recombinant DNA technology have made it possible to generate forms of recombinant antibody fragments which, in many ways, are functional substitutes of larger intact immunoglobulin molecules. Single heavy domain ("dAb") antibody fragments have been the subject of several reports in the patent and scientific literature. The literature reports efforts to generate phage display libraries of such fragments for biopanning against a target ligand.

U.S. Patent No. 5,702,892 ('892) discloses a phage display library constructed in an M13 derived expression vector, in which recombinant phage of the library contain a polynucleotide encoding a fusion protein which comprises a phage coat protein and an

immunoglobulin heavy chain binding-fragment. The heavy-chain binding-fragment spans from a position upstream of CDR1 to a position downstream of CDR3. '892 describes that the DNA sequence encoding the CDR3 region and/or the CDR1 region may be randomly varied so that the population of phage expresses a series of potential heavy chain binding domains for panning against the target ligand.

U.S. Patent No. 5,759,808 discloses a phage display library comprising a population of phage based on random variation of a cDNA sequence obtained from lymphocytes of camelids previously immunized with target antigens. Camelid heavy chain antibodies occur naturally, in a composition of about 45%, as heavy chain dimers. Heavy chain antibodies specific for a target antigen may be generated by immunizing a member of the camelid species with the target antigen (see Lauwereys et al. (1998) The EMBO J. 17, 3512-3520).

Hamers-Casterman et al. (1993) Nature 363, 446-448 report that camelid heavy chain antibodies are naturally more hydrophilic at amino acid residues at locations 44, 45 and 47 (Kabat numbering system), in FR2, which corresponds to the surface where they normally contact the V_L domain. Another salient feature of a camelid V_H is that it generally has a comparatively longer CDR3 with a high incidence of cysteines and thus may form, via paired cysteines in CDR1 and CDR3, exposed loops, which are more amenable to binding into cavities such as the active site of enzymes and antibodies (Desmyter et al. (1996) Nat. Struct. Biol. Vol. 3, No. 9, p. 803). However, it has been questioned whether single domain antibodies with desired affinities can be generated with such configurations in the absence of prior immunization, i.e. with a naïve library (Lauwereys et al. (1998) supra).

The present invention discloses advances in the technology related to creating libraries containing immunoglobulin-like proteins that specifically bind target ligands eg. antigens.

Summary of the Invention

The invention is directed to a population of variants of at least one parental ligand-binding molecule, wherein said parental ligand-binding molecule comprises an immunoglobulin V_H binding fragment comprising, at least in substantial part, at least the framework (FR) regions of the immunoglobulin V_H fragment depicted in one of Figures 1 or 2 and wherein said variants comprise at least in substantial part, the FR regions of the immunoglobulin V_H fragment depicted in one of Figures 1 or 2 and differ from said parental ligand-binding molecule in amino acid residues constituting at least part of at least one of the CDRs of said parental ligand-binding molecule. Preferably said population of variants is constituted by one or more combinatorial libraries of such variants, for example, protein arrays, phage display libraries, ribosome display libraries etc.

It is to be understood that the variants may (though not necessarily) form part of another structure or molecule, for example in the case of phage display, part of the coat protein of the phage. Accordingly, the term variant is used broadly to refer to variants of the essential molecule (a ligand-binding molecule) when forming part of another structure or molecule (eg. as in phage display or ribosome display) or when independent of any such combination, eg. in the case of protein arrays whose members maynot be associated with individual supporting structures/molecules.

In another aspect, the invention is directed to a ligand-binding molecule which has been identified as binding to a target ligand by screening a combinatorial library of the invention for one or more ligand-binding molecules which specifically recognize said target ligand. The invention is also directed more generally to any specific such ligand-binding molecule which is derived from such combinatorial library of the invention. It is understood herein that such specific ligand-binding molecule may be directly obtained from such a library or may be indirectly derived, for example, through the course of further antibody engineering or other modification steps (eg. creating fragments, derivatives, a secondary library, etc) using a ligand-binding molecule directly or indirectly obtained from such library. It also understood that the invention excludes known ligand-binding molecules. In one embodiment of this aspect of the invention the target ligand is a cancer antigen.

Figures 2, 3 and 4 depict preferred variations, more fully described below, on the preferred immunoglobulin V_H binding fragment and/or nucleic acid construct (depicted in Figure 1. Figure 1 describes a wild-type parental immunoglobulin V_H binding fragment derived from human monoclonal antibody BT32/A6 (hereinafter referred to as "A6") partially described in U.S Patent No. 5,639,863 (hereinafter referred to as the '863 patent). It has now been found that A6 has preferred solubility and other characteristics which lend themselves well to the creation of libraries, including naïve libraries, of various types of human immunoglobulin fragments including scFvs, Fds, Fabs etc., as more fully described below. Accordingly, A6, and in particular, as more fully described below, at least a substantial part of the framework (FR) regions of the A6 V_H fragment depicted in Figure 1, alone or in combination with features of its CDR3, provide a useful departure point, in the form of a **parental** ligand-binding molecule, for the randomization or partial randomization of amino acid residues

which tend to play a predominant role in ligand-binding, namely the CDR regions of the heavy chain and particularly the CDR3 of the heavy chain. As more fully described below, the nucleic acid changes (removal of the recombination site) relative to A6 wild-type Figure 1reflected in Figure 14 may be incorporated into Figures 1 or 2 to create preferred Figures 3 and 4, respectively.

The combinatorial library of the invention may be generated by phage display. Accordingly, in a preferred embodiment of one aspect of the invention, the invention is directed to a phage display library displaying a plurality of different variants of a parental ligand-binding molecule, wherein said parental ligand-binding molecule comprises an immunoglobulin V_H binding fragment comprising, at least in substantial part, at least the FR regions of the immunoglobulin V_H fragment depicted in one of Figures 1 or 2 and wherein said variants are encoded by nucleic acid sequences which vary from the nucleic acid sequence encoding said parental ligand-binding molecule in a sub-sequence (at least one) encoding at least part of one of the CDRs of said parental ligand-binding molecule, preferably the CDR3, whereby said plurality of variants comprise at least, in substantial part, the FR regions of the immunoglobulin V_H fragment depicted in one of Figures 1 or 2 and are differentiated, at least in part, by amino acid variations encoded by variations in said sub-sequence.

In a preferred embodiment, in addition to substantial preservation and optional improvement of the FR regions of A6, the A6-based parental ligand molecule comprises (and therefore preserves within members of the library), in substantial part (subject to at least partial randomization of selected regions of one of the CDRs, preferably the CDR3, to create

binding diversity within the library), one or more of the CDR regions of the A6 V_H fragment, particularly the CDR3. In a further preferred embodiment, at least the length of the wild-type V_H CDR3 (23 amino acids) and preferably also elements of its amino acid composition, is preserved or at least partially preserved (approximately16-23 amino acids and more particularly 18 to 23 amino acids). Optionally the CDR3 may also be lengthened by approximately 1 to 10 residues. The library may optionally have representation of binding molecules having CDR3s of varying lengths.

In a preferred aspect of the invention the parental ligand-binding molecule is a dAb fragment. It is known that a dAb molecule , due to the removal of its light chain partner, tends to, in most, if not all cases, aggregate, in varying degrees due to the "sticky" nature of the V_L interface. This stickyness is attributable, at least in part, to the hydrophobic nature of the V_H residues at this interface. This stickyness results in substantial dimer and/or multimer formation which may reduce, on the whole, the solubility characteristics of members of the library. Accordingly, in a further preferred aspect of the invention A6 V_H amino acid residues at the VL interface are substituted by residues which tend to minimize aggregate formation, for example, hydrophilic amino acids, and preferably one or more of the substitutions reflected in Figure 2, relative to Figure 1.

Alternatively, in yet a further preferred embodiment of the invention, more fully described below, such substitutions are not fixed within the entire population of the library, but are introduced by randomizing or partially randomizing various A6 V_H amino acid residues, particularly including FR residues, among the residues at the interface. (see for example,

Padlan et al "Anatomy of the Antibody Molecule" Molecular Immunology Vol. 31, p169-217, Table 25 for itemization and related discussion of these residues).

Alternatively, in yet a further preferred embodiment of the invention, FR regions, other than, or in addition to, modifications to the V_L interface (FR2) may be modified by at least partial randomization, for example, one or both of FR1 (one or more of residues 4 to 21) and FR4 (one or more of residues 100o to 113) to improve, on the whole, the solubility characteristics of members of the library (for example, biasing at least some and preferably all of one or both of these sets of residues (at least 70% or more), preferably 90% in favour of the parental amino acid constitution to achieve 10% randomization).

In the case of A6 dAb fragments, it has been found that recombination events within the nucleic acid sequence encoding the V_H binding fragment tend to result in deletions yielding shorter molecules, with possibly comprimised binding characteristics. Thus, in a further preferred aspect of the invention exemplified in Figures 3 and 4, and nucleic acid sequences which promote such recombination events (at putative recombination sites) are substituted, to oppose this tendency, preferably in a manner that does not result in an amino acid change. These changes may be incorporated into the wild-type A6 (see Figure 3) or improved variations thereof exemplified in Figure 2 (see Figure 4) which have a reduced tendency towards aggregation.

Thus, in particularly preferred aspects, the present invention provides a heterogeneous population of genetic packages (eg. phage) having a genetically determined outer surface protein, wherein the genetic packages collectively display a plurality of different, preferably **human, (ie. having substantial identity, preferably at least 80% homology to human**

framework and other conserved regions) V_H ligand-binding fragments, each genetic package including a nucleic acid construct coding for a fusion protein which comprises at least a portion of the outer surface protein and a variant of at least one soluble parental ligand-binding fragment preferably derived from or having a substantial part of the FR regions of the amino acid sequence identified in one of Figures 1 or 2 (or a sequence at least 80%, preferably 85 to 100%, more preferably 90-100%, homologous (% identity) thereto), wherein the variant V_H ligand-binding fragments preferably span from a position upstream of an immunoglobulin heavy chain CDR1 to a position downstream of CDR3 (preferably including substantially all of FR1 and/or FR4), and wherein at least part of a CDR, preferably the CDR3, is a randomly generated variant of a CDR of said parental V_H ligand binding-fragment and wherein the fusion protein is preferably expressed in the absence of an immunoglobulin light chain whereby the variant V_H ligand-binding fragments are, on the whole, better adapted to be or better capable of being expressed as soluble proteins.

In yet another embodiment of the invention, by biasing the amino acid constitution, preferably on an individual amino acid by amino acid basis, in favor of the wild-type or parental amino acid constitution, even portions of the parental ligand-binding molecule that are randomized in favor of generating variability in the variant binding fragments can be engineered to maintain favorable solubility characteristics of the parental binding domain. Preferably, a portion of the construct encoding at least part of the CDR3 is biased or partially biased in favor of the parental amino acid constitution.

In a further preferred embodiment, the parental V_H binding-fragment naturally has a long CDR3 that is amenable to forming exposed loops for binding into cavities. In a most

preferred embodiment, the parental V_H ligand-binding fragment is built on a human framework or is adapted from or adaptable to a human framework.

In another preferred embodiment, the preferred binding region of the variants (corresponding to the randomized or partially randomized part of the CDR3) is located in carboxy terminal region of the CDR3.

In summary, according to the invention, a substantial part of the amino acid sequence identified in Figure 1, preferably including at least part of the CDR3, supplies the preferred amino acid constitution of the various preferred parental ligand-binding molecules, such that a population of variant heavy chain ligand-binding molecules built on this framework of amino acids are on the whole better adapted to be or better capable of being expressed as soluble proteins.

Brief Description of the Drawings

The invention will now be described with reference to the drawings, wherein:

Figure 1 is a sequence diagram showing a parental V_H ligand-binding molecule (A6) according to the invention.

Figure 2 is a sequence diagram showing another parental V_H ligand-binding molecule (A6.1) according to the invention, additionally showing modified nucleic acid bases corresponding

to amino acids 24 and 25, for introducing the NheI site. Introduction of this NheI site does not alter the amino acid constitution of A6.1.

Figure 3 is a sequence diagram showing the A6 V_H ligand-binding molecule (encoded by A6-chi(-)) according to the invention, in which the nucleic acid residues corresponding to amino acids 3 to 16 of A6 wild-type have been modified to remove a putative recombination site, leaving the amino acid constitution of A6 unchanged.

Figure 4 is a sequence diagram showing the A6.1 V_H ligand-binding molecule (encoded by A6.1-chi (-)) according to the invention, in which nucleic acids corresponding to amino acids 3 to 16 of A6.1 have been modified to remove a putative recombination site, leaving the amino acid constitution of A6.1 unchanged. The altered nucleic acid residues corresponding to NheI are also shown.

Figure 5 is a facsimile of an SDS-PAGE showing high expression of human A6.1 dAb in E. Coli.

Figure 6 shows size exclusion chromatograms of molecular weight markers (A), dAb (B) and A6.1 dAb obtained by gel filtration obtained using a Superdex 75 column. The masses of the markers were 2,000, 67, 43, 25 and 14 kDa.

Figure 7 is a size exclusion chromatogram of A6.1 dAb following IMAC purification showing molecular weights associated with the peaks.

Figure 8 is an NMR 2-D spectra showing the molecular configuration of two embodiments of an A6.1 based dAb. In particular, the two 15N-1H HSQC spectra are: a: R3A10(Cys-), the spectrum was acquired at 308 K; b: M2R2-1, the spectrum was acquired at 298 K.

Figure 9 is diagrammatic representation of the amino acid substitutions in parental ligand-binding molecule A6.1 and A6.1C relative to wild-type A6.

Figure 10 is a sequence diagram showing a parental V_H ligand-binding molecule designated A6.1C.

Figure 11 is a graphic representation of the binding characteristics of A6.1C library binders to 3B1 and control BSA.

Figure 12 is a sensogram overlay showing the binding characteristics of a potential V_H binding fragment generated against anti-FLAG antibody (M2) using a phage display library of the invention.

Figure 13 is a diagrammatic representation of vector, SJFI, used to create the vector into which the library is cloned.

Figure 14 is a listing of the nucleotide and amino acid sequence of A6 V_H after introduction of the *Nhe*1 site and removal of the putative recombination site at amino acid residues 3 to 16.

Figure 15 is a schematic representation of steps taken to remove the putative recombination site of the 5 end of the A6 V_H gene.

SEQ. ID. NO. 1 corresponds to the nucleic acid sequence shown in Figure 1.

SEQ. ID. NO. 2 corresponds to the amino acid sequence shown in Figure 1.

SEQ. ID. NO. 3 corresponds to the amino acid sequence of CDR1 shown in Figure 1.

SEQ. ID. NO. 4 corresponds to the amino acid sequence of CDR2 shown in Figure 1.

SEQ. ID. NO. 5 corresponds to the amino acid sequence of CDR3 shown in Figure 1.

SEQ. ID. NOS. 6-11 and 25 correspond to the nucleotide sequences of primers disclosed herein.

SEQ. ID. NOS. 12-24 correspond to the amino acid sequences of CDR3 variants disclosed herein at Table 2.

Detailed Description of Preferred Embodiments

In a preferred embodiment, the invention is directed to a population of genetic packages having a genetically determined outer surface protein including genetic packages which collectively display a plurality of different ligand-binding molecules in association with the outer surface protein, each package including a nucleic acid construct coding for a fusion

protein which is at least a portion of the outer surface protein and a variant of at least one soluble parental ligand-binding molecule derived from or having the amino acid sequence identified in Figure 1 (or a sequence preferably at least 80% homologous in the framework and conserved regions thereof), wherein at least part of the construct, preferably including at least part of the CDR3 identified in Figure 1, encodes or is biased in favor of encoding, the amino acid constitution of the parental ligand binding fragment such that the plurality of different ligand-binding domains are on the whole better adapted to be or better capable of being expressed as soluble proteins. The variant V_H ligand-binding molecules are preferably characterized by a CDR3 having 16 to 33 amino acids.

Preferably, the replicable genetic package is a recombinant phage and the heterogeneous population of replicable genetic packages collectively constitute a phage display library.

In a preferred embodiment, the parental ligand-binding molecule is a V_H binding fragment, and the plurality of variant ligand-binding fragments are expressed in the absence of light chains. In another embodiment, the parental ligand-binding-molecule is a natural occurring antibody or fragment thereof, having a natural human V_L interface. In another embodiment, the V_L interface is engineered to avoid hydrophobic amino acids. In another embodiment, the V_L interface is engineered for amino acids, which form weak interactions. In another embodiment the parental ligand binding molecule has a camelid type V_L interface. In another embodiment, at least one of the V_L interface amino acids are randomized or partially randomized in the construction of the library.

Preferably the potential V_H binding fragments include the entire FR1 through to FR4 regions, although it is to be understood that partial deletions, for example, within CDR2, are contemplated to be within the scope of the invention.

Preferably, CDR3s of a variety of different lengths from 16 to 33 amino acids are predominantly represented among the potential V_H binding fragments. Preferably CDR3s of a variety of different lengths, from 18 to 28 amino acids, or from 20 to 25, or from 18 to 23, amino acids are predominantly represented in the library. In a preferred embodiment of the invention, the parental V_H ligand-binding fragment is built on a human framework and preferably is the parental V_H ligand-binding fragment identified in Figure 1 which has a CDR3 of 23 amino acids in length.

The invention encompasses a phage display library which is constructed using a parental V_H ligand-binding molecule derived from a human parental V_H ligand-binding fragment, or is built on any framework which is at least 80% (preferably 85%, more preferably 90 to 95%) homologous to the framework and other conserved regions of a fully human V_H chain. The invention also contemplates that the parental V_H binding-fragment, though not human, is adapted (eg. humanized) or adaptable (eg. to be adapted after selection of preferred binders) to a human framework.

In another embodiment, the invention also contemplates the random, biased or fixed occurrence of features disclosed in the camelid literature, for example pairable cysteines in CDR1 and CDR3 (optional) and/or the substitution of hydrophilic amino acids at least one of positions 44, 45, and 47 and preferably also positions 93 and 94 (Kabat numbering system).

In a most preferred embodiment of the invention, the parental ligand-binding molecule is a V_H fragment derived from a human IgM heavy chain, and preferably comprises FR1 through FR4 of the V_H chain. A partial sequence of the preferred antibody BT32/A6 (A6) is disclosed in U.S. Patent No. 5,639,863, incorporated herein by reference. The entire sequence is supplied now in Figure 1.

In Figure 1, the CDR regions are demarcated. The amino acid residue numbers in Figure 1 and throughout the disclosure refer to the Kabat numbering system (Kabat et al. 1991, Sequences of Proteins of Immunological Interest, publication No. 91-3242, U.S. Public Health Services, NIH, Bethesda MD) except in the sequence listings and where explicitly stated or otherwise implied. Figure 1 corresponds to SEQ. ID. NO. 1 (nucleic acid) and SEQ.ID. NO. 2 (amino acid). Figure 1 demarcates and labels regions CDR1 (corresponding to SEQ. ID. NO. 3), CDR2 (SEQ. ID. NO. 4) and CDR3 (SEQ. ID. NO. 5).

In addition to other types of antibody fragments (e.g. scFv, FAb, etc.) the A6 framework provides preferred solubility characteristics for creating dAb libraries. The term preferred solubility characteristics, as used herein, refers to at least one of the several, often correlated, characteristics including good yield, expression as a soluble product (as opposed to inclusion bodies) within the periplasm of the host organism, eg. *Escherichia. Coli,* and a reduced tendency to dimerize and other aggregate formation.

The terms "polypeptide", "peptide" and "protein", unless the context implies otherwise, are used interchangeably herein, to refer to polymers of amino acid residues of any length.

The term "combinatorial library" is used herein to refer to a set of molecules, typically belonging to a defined (narrowly or broadly) class comprising a substantial number of potentially useful variants, wherein the variations in the molecule represent a complete or partial set of permutations or combinations of at least some constituent elements of a reference molecule, which is typically a template or "parental" molecule, or simply the class itself. For clarity, in the case of polypeptides and nucleic acids, the constituent elements are amino acids and nucleic acid bases, respectively.

As used herein, the phrase **"in substantial part"** refers to variations relative to a referenced molecule which do not significantly impair the "functionality" of that molecule. In the case of the parental ligand-binding molecule and variants thereof, functionality refers primarily to the solubility and binding characteristics of the molecule. Such variations (ie. the referenced molecule in substantial part) can be tested systematically to assess their impact. In the case of framework regions, in contrast to CDR regions, due to the substantial conservation of the framework amino acid residues, a substantial part of the framework would preferably refer to at least 80% identity of the amino acid residues and more preferably an 85 to 100% identity, and even more preferably at least a 90% identity of the amino acid residues. However, it is understood that each of the previous percentages could be relaxed to discount instances where the absence of identity in a given residue, is due to a well recognized conservative amino acid substitution, or where a particular class of functionality is noted, e.g. hydrophilic, if the substitution is with a residue of the same class. In the case of CDR residues, these numbers could be considerably even more relaxed. The term "in substantial part", in reference to portions of framework and CDR regions, also contemplates the possibility of

additions and deletions in those regions which do not impact the solubility and binding characteristics of the ligand-binding molecule in question.

The term ligand-binding fragment is used broadly to define the whole or any part of an antibody that is capable of specifically binding to any ligand, in the broadest sense of the term ligand.

An A6-based human heavy domain ligand-binding-fragment is well suited for the development of a combinatorial library (optionally a phage display library) that is used to generate soluble binding fragments that are useful for human diagnosis and therapy (due to limited HAMA response). These phage display libraries are used to selectively generate molecular probes that specifically interact with a ligand, including without limitation, natural and synthetic molecules and macromolecules and can be used *in vitro* (i.e., a diagnostic) and *in vivo* (i.e., a diagnostic and/or therapeutic) as indicators, inhibitors and immunological agents. The types of natural and synthetic molecules and macromolecules include but are not limited to: antibodies and fragments thereof; enzymes; cell receptors; proteins, polypeptides, peptides; polynucleotides, oligonucleotides; carbohydrates such as polysaccharides, oligosaccharides, saccharides; lipids; organic-based and inorganic-based molecules such as antibiotics, steroids, hormones, pesticides, herbicides, dyes, polymers.

As shown in Figure 5, a facsimile of an SDS-Page, A6.1 V_H has preferred solubility characteristics. This SDS-PAGE shows a particularly heavy band **showing strong expression in E. Coli of an A6.1 dAb, designated R3A10.** R3A10 was expressed as a soluble V_H in E. Coli. Yields as high as 55 mg/L of bacterial culture were obtained by IMAC

chromatography of periplasmic extracts. The single domain product was shown to be highly pure and homogeneous by SDS-PAGE (Figure 5). Size exclusion chromatography on a Superdex 75 column gave a symmetric single peak at the expected elution position of a monomeric molecule with a molecular weight of 16 kDa, the molecular weight of V_H (Figure 7). A preparation of R3A10 gave very high quality NMR data in the absence of detergent, confirming the absence of aggregated material (see Figure 8A).

In general, the protein yields of many dAbs from the A6.1 library were above 5 mg per liter of bacterial culture in shaker flasks. Some had yields more than 10 mg and one over 50 mg. The solubility of the wild type and the camelized versions were very high as shown by NMR studies. R3A10 and M2R2-1(Cys^-) for example, were soluble in mM concentrations over extended periods of time allowing good quality NMR data collection. A NMR structure of a human VH camelized in this manner has been described (Reichmann, J Mol, Biol) but in order to reduce aggregation and achieve sufficient solubility CHAPS detergent had to be added to the sample during NMR data collection. By contrast the A6.1 dAb molecules described here were completely free of aggregated material in the absence of detergent.

Conventional antibodies such as those found in human or murine species are composed of two identical light chains and two identical heavy chains. The combining sites of these antibodies are formed by association of the variable domains of both chains. This association is mediated through hydrophobic interactions at the interface. Structural and biochemical studies have shown that the heavy chain variable domain (V_H) provides most of the antigen-contacting residues (Padlan, 1994) (Chothia & Lesk, 1987) (Chothia, Novotny, et al., 1985). This finding has formed the basis for the development of single heavy domain

antibodies (dAbs) - recombinant antigen binding fragments consisting of only the V_H (Ward, Gussow, et al., 1989) (Cai & Garen, 1996). However, in the absence of their V_L partners, V_Hs have been found to be insoluble, presumably because of the exposed hydrophobic V_L interface (Ward, Gussow, et al., 1989). Heavy chain antibodies, found in camelids (Hamers, Atarhouch, et al., 1993) (Sheriff & Constantine,), lack light chains and as a result have variable domains that reflect the absence of a V_L partner. Single domain antibodies derived from these antibodies are highly soluble and the structural basis of solubility has been partially elucidated. First, conserved human/murine interface residues such as Val37, Gly44, Leu45 and Trp47 are generally replaced in heavy chain antibodies by tyrosine or phenylalanine, glutamate, arginine or cysteine, and glycine, respectively. These mutations increase the hydrophilicity of the V_L interface either by non-polar to polar substitutions or, in a more subtle way, by inducing local conformational changes (Desmyter, Transue, et al., 1996) (Spinelli, Frenken, et al.,). This explanation is supported by experiments in which an insoluble human V_H was made soluble by introducing the aforementioned mutations at positions 44, 45 and 47 (Davies & Riechmann, 1994). Second, in the solved structures of two camel dAbs, the CDR3s fold back on the V_H surface, masking a significant surface area of the V_L interface (Desmyter, Transue, et al., 1996)(Decanniere, Desmyter, et al., 1999).

Several other features of V_HHs are noteworthy. One is the frequent occurrence of the cysteine residues in CDR1 and CDR3 (Muyldermans, Atarhouch, et al., 1994) (Lauwereys, Arbabi, et al., 1998 (Vu, Ghahroudi, et al., 1997). While the location of the CDR1 cysteine is typically fixed at position 33, that of the CDR3 cysteine varies. These two residues form a disulfide linkage between CDR1 and CDR3 (Desmyter, Transue, et al., 1996) (Davies & Riechmann, 1996). In the crystal structure of a dAb-lysozyme complex, the disulfide linkage imparts rigidity on the CDR3 loop which extends out of the combining site and penetrates

deep into the active site of lysozyme (Desmyter, Transue, et al., 1996). A second feature is the longer average length of the V_HH CDR3, relative to human or murine V_Hs (Muyldermans, Atarhouch, et al., 1994). A longer CDR3, which is a feature of A6, increases the antigen binding surface and, to some extent, compensates for the absence of the antigen binding surface provided by the V_L in conventional antibodies (Desmyter, Transue, et al., 1996). A third feature is the absence of the CDR3 salt linkage that is typically present in conventional antibodies and formed by arginine or lysine residues at position 94 and aspartate at position 101 (Desmyter, Transue, et al., 1996) (Muyldermans, Atarhouch, et al., 1994) (Spinelli, Frenken, et al., 1996) (Davies & Riechmann, 1996) (Chothia & Lesk, 1987) (Morea, Tramontano, et al., 1998).

As antigen binding fragments, dAbs are an attractive alternative to scFvs because of their much smaller size and the fact that they demonstrate affinities comparable to those demonstrated by scFvs (Ward, Gussow, et al., 1989) (Spinelli, Frenken, et al., 1996) (Lauwereys, Arbabi, et al., 1998) (Davies & Riechmann, 1995) (Arbabi, Desmyter, et al., 1997) (Reiter, Schuck, et al., 1999). Smaller size is an advantage in applications requiring tissue penetration and rapid blood clearance. Smaller molecules also offer a tremendous advantage in terms of structural studies (Davies & Riechmann, 1994) (Constantine, Goldfarb, et al., 1992 (Constantine, Goldfarb, et al., 1993).

Phage antibody library construction is much simpler and more efficient if single domain antibodies are used instead of Fabs or single chain Fvs. Randomization can be introduced at a much higher percentage of CDR positions without exceeding practical library size. The problem of shuffling original V_L-V_H pairings is also avoided. Camelid phage dAb libraries constructed from the V_HH repertoire of camels immunized with target antigens have performed well (Arbabi, Desmyter, et al., 1997) (Lauwereys, Arbabi, et al., 1998)

(Decanniere, Desmyter, et al., 1999). However, construction of libraries from immunized camels presents obvious problems. In addition, the non-human nature of products from these libraries limits their usefulness. Synthetic dAb libraries (Davies & Riechmann, 1995) (Reiter, Schuck, et al., 1999), particularly those based on a human V_H framework, alleviate these problems.

Thus according to another embodiment of the invention, the parental ligand-binding fragment has amino acid substitutions at V_L interface which reduce the tendency to aggregation attributable to the "stickyness" of the V_H dAb at this interface. In another preferred embodiment of the invention, the parental ligand-binding fragment has a long CDR3 similar to some camelid antibodies. As discussed above, according to another embodiment of the invention, an A6 dAb based library is preferred, because A6 has an unusually long CDR3 of 23 amino acids. In a particularly preferred embodiment, the library preserves the entire length of this CDR3 and at least one of positions 44, 45 and 47 are altered, preferably 44 or 45 to camelid type residues. In the embodiment exemplified in examples 5 and 6, the CDR3 was randomized and cysteine residues were introduced at positions 33 and 100e in the expectation that the residues would form the CDR1-CDR3 disulphide bridge present in the camel antibody Cab-Lys3 (Desmyter, Transue, et al. 1996). The library was evaluated by panning against an IgG that binds a peptide of known sequence. Procedures for the construction and testing of this library is described in examples 5 through 13 and below as follows.

A6 dAb is expressed surprisingly well as a soluble product in *E. coli* with a yield of approximately 10 mg per liter of bacterial culture . Mass spectrometry has confirmed that the

product has the expected molecular weight. As shown in Figure 6, size exclusion chromatography of the product reveals three components that are thought to correspond to monomer, dimer and higher oligomer on the basis of their elution volumes (Figure 6A and B). However, the monomer peak elutes unusually late suggesting that the dAb is interacting non-specifically with the gel matrix possibly through its exposed hydrophobic V_L interface. This is a property of human and murine-derived dAb that is not unusual and which has been documented previously (Ward, Gussow, et al., 1989) (Davies & Riechmann, 1994). By introducing the Gly44Glu, Leu45Arg, and Tyr47Gly mutations (Davies & Riechmann, 1994) into the A6 framework, a product is obtained that is exclusively monomer and which elutes at the expected volume in size exclusion chromatograms (Figure 6C).

To generate the template for library construction, the dAb was further modified by introducing Val93Ala and Lys94Ala mutations in FR3 and an Ala33Cys mutation in CDR1. A preferred library having the itemized substitutions at positions 44, 45, 47 93 and 94 is referred to as "A6.1" Libraries with Ala 33 Cys and Cys at position 100e (see discussion below) are termed "A6.1C". These substitutions are diagrammatically illustrated in Figure 9 and the sequences for A6.1 and A6.1C are shown in Figures 2 and 10, respectively. In camelid V_HHs, positions 93 and 94 are predominantly occupied by Ala residues and Cys is frequently found at position 33 (Muyldermans, Atarhouch, et al., 1994) (Vu, Ghahroudi, et al., 1997). The library was constructed by randomizing 19 amino acids in CDR3, leaving the last three residues, Phe100, Asp101, and Tyr102, unchanged (Davies & Riechmann, 1995). The degenerate oligonucleotide used for randomization was designed so as to always introduce Cys at position 100e. This was done to facilitate the formation of intra-molecular disulfide linkage between 100eCys and 33Cys in CDR3 and CDR1, respectively

(Muyldermans, Atarhouch, et al., 1994) (Desmyter, Transue, et al., 1996) (Davies & Riechmann, 1996).

The library was initially placed in a phagemid vector. Following transformation the size of the library was determined to be 2:1 x 10^7. Of 80 randomly picked clones analyzed by PCR all but one had the dAb insert. In addition, all twenty that were sequenced were unique, demonstrating the diversity of the library. To convert the display format from monovalent to multivalent, the library was sub-cloned into a phage vector (MacKenzie and To, (1998). Following transformation, the size of the library was determined to be 6.6 x 10^7. Therefore, on a random basis each member of the original phagemid library is represented 3 times.

Initially, the library was panned, in both formats, against 3B1 scFv, which is specific for a bacterial carbohydrate (Deng, MacKenzie, et al., 1994). With the phagemid vector format, panning failed to enrich for binders and PCR analysis of clones selected at different stages of the panning process revealed almost universal deletion of the dAb inserts. However, with the phage vector seven different dAbs that bound to 3B1 were identified. As shown in Figure 11, these dAbs bound to the target antigen, 3B1, in ELISA experiments and showed no detectable binding to the control BSA. In each instance, the consensus sequence was present at the extreme C-terminal end of CDR3 (see Table 1 and Table 2 below).

Table 1. The CDR3 sequences of dAbs isolated by panning against 3B1 scFv. The consensus sequence is shown in bold.

dAb	CDR3 sequence
3B1R3-1	VGPITGGAPRAVCKH**AKAWFLP**FDI
3B1R2-3	SSQPRVTSSPCVA**SKSWFLP**FDI
3B1R2-2	PTTGIRGEKDCTP**KKMWRLP**FDI
3B1R2-4	RDPSVTDTGCCTP**RWQAWLP**FDI
3B1R3-3	PGEPPEASAPCLR**HRVGWLP**FDI
3B1R3-15	KTVKMRDDEVCTK**RTNWLLP**FDI
3B1R3-19	PGNVASQQNLCGL**RATRWLP**FDI

In the phage vector format, the library was also panned against M2 IgG, an antibody which was raised against the FLAG peptide DYKDDDDK (Knappik & Pluckthun, 1994). More recent studies showed that M2 recognizes the consensus sequence XYKXXD and prefers epitopes with aspartate at the first position (Miceli, Degraaf, et al., 1994). Twenty-four different dAbs with the FLAG consensus sequence were identified the sequencing of clones randomly selected after 3 rounds of panning (Table 2 below).

Table 2. The CDR3 sequences of dAbs isolated by panning against M2 IgG at different DTT concentrations. The FLAG consensus sequence is shown in bold.

A. No DTT	
VQYGKHRRGSCIEVHP**EYKDFDI**[a]	**C. 1 mM DTT**

NPPKPGAQARCVTTVK**DYKEFDI**[b]	TAEPALSPQACMTKER**QYKDFD**I
AAIQTETARWCDRHPV SYKMFDI[c]	QTETQPLYNDCILRQA**GYKWFD**I
QTETQPLYNDCILRQA**GYKWFDI**[d]	AAIQTETARWCDRHPV**SYKMFD**I
MHTLQHYRNLCSYQLA**DYKHFDI**[e]	AADPRALMKSCALVTS**DYKWFD**I
GLSGSRPNEQC**DYKTGD**HVQFDI[f]	LRGRMRQQSCCGGAGN**TYKDFD**I
LSGQNYTKTRCLVMQN**DYKMFDI**[g]	NPPKPGAQARCVTTVK**DYKEFD**I
TAEPALSPQACMTKER**QYKDFDI**[h]	PGPGAPG**EYKCED**WSNRQLSFDI
ETYMYTRGKYCRALSA**DYKLFDI**[i]	GGLKNQ**DYKRCD**TEGSGFTRFDI
ESKASRTADQCSGPTP**GYKNFDI**[j]	
GSQAIKNLSECLVRSD**DYKKFDI**[k]	**D.10 mM DTT**
GRYFQSKITSCENNDR**DYKLFDI**[l]	DRGPQGAPDPCLQIIN**DYKTFD**I
PRPARTGHKTCFVRPK**NYKDFD**I	YMPAASPVSQCLATLI**EYKAFD**I
AEAHSQLPPRCRRKTD**EYKIFD**I	KHEVNHVEDRCNQTTE**TYKMFD**I
SHKTSQPVRNCSATDN**SYKLFD**I	LQNSPKNSGWCDFILA**GYKAFD**I
TMGTLHSPHECMKSLV**TYKNFD**I	MEPNRSYRGLCLEAPN**EYKWFD**I
GRYFQSKITSCENNDR**DYKLFD**I	GDKQSPKSRRCLTWLV**GYKHFD**I
ELGWRPRVQACHYSRN**DYKYFD**I	PEQRQTVGHVCLTRMP**DYKHFD**I
KDVTRTNTVSCSKDRQ**DYKMFD**I	QHTWARQENGCFMV**DYKFSD**FDI
YSATAKWRDKCYEKSR**DYKMFD**I	
YEIVPFIASRCVIERA**DYKLFD**I	**E.100 mM DTT**
ADAPNRQKERCVVAVH**GYKRFD**I	EAESRTWYAPCHSTRT**DYKLFD**I
NEEKFSVYSECELYLP**TYKMFD**I	TAEPALSPQACMTKER**QYKDFD**I
IWEGEKHYAECVTG**TKYQPD**FDI	KDVTRTNTVSCSKDRQ**DYKMFD**I
	NWDAKDSPRKCSLMLT**MYKDFD**I

B. 1 μM DTT	GRLNHRSQTTCLVSEK**EYKSFD**I
NPPKPGAQARCVTTVK**DYKEFD**I	GAQYRRLTSSCKPRSH**EYKEFD**I
ETPRDTKLTACKFMPS**DYKYFD**I	RNGNLTYKASCSSAGD**DYRDFD**I
AAIQTETARWCDRHPV**SYKMFD**I	CDQTYCSKWLCREEQV**DYKLFD**I
TAEPALSPQACMTKER**QYKDFD**I	PNERIKEGANCSMGTT**EYKQFD**I
LSGQNYTKTRCLVMQN**DYKMFD**I	PRSLSTSNATCTTS**DYKRHD**FDI
NLPQPLRERTCIGPRR**DYKMFD**I	
SVPRITDIQTCQTLHS**DYKHFD**I	
DRALGLNDTWCRGPRM**SYKWFD**I	
MHTLQHYRNLCSYQLA**DYKHFD**I	
LPQFIPNHMLCNYQSV**DYKTFD**I	
QDWHWQEQRSCPVTDF**RYKDFD**I	
RANEYGSKSRCTEGMY**EYKSFD**I	
GAMPQGASRMCAADQR**EYKAFD**I	

[a]M2R2-1; [b]M2R2-2; [c]M2R2-4; [d]M2R2-5; [e]M2R2-9; [f]M2R2-10; [g]M2R2-13; [h]M2R2-14; [i]M2R2-15; [j]M2R2-18; [k]M2R3-4; [l]M2R3-13.

No consensus sequence other than XYKXXD could be identified. Interestingly, like the 3B1 binders, all the FLAG consensus sequences occurred in the C-terminal half of CDR3 and with two exceptions all occupied identical positions. To ascertain if this observation was related to the presence of CDR1-CDR3 disulfide linkage, the reduced version of the same library was also panned against M2 IgG. This was done by the addition of an appropriate concentration of DTT to the phage mixture during the binding stage of the panning procedure. Panning was performed at 1 M, 1 M, 10 M and 100 M DTT. The same

concentration of DTT was also included in the wash buffer to maintain the phage dAbs in the reduced state. The CDR3 sequences of the dAbs thus isolated are given in Table 2. As in the absence of DTT and at all DTT concentrations, the FLAG consensus sequences were located near the C-terminal end of CDR3.

The binding (binding kinetics) to M-2 IgG of five of the dAbs listed in Table 2 (M2R2-2, M2R2-4, M2R2-10, M2R2-13 and M2R3-4 were investigated by surface plasmon resonance. It was observed that the binding data fit poorly to a 1:1 interaction model in all instances, making the derivation of kinetic and affinity constants impossible. When binding studies were conducted in the presence of DTT it was observed that the amount of binding increased significantly, particularly for M2R2-2. Furthermore, data collected in the presence of DTT fit much better to a 1:1 interaction model. In view of this result a M2R2 mutant lacking the CDR1-CDR3 disulphide bridge was constructed and expressed for BIACORE studies. The data for the binding of this mutant to immobilized M2R2-2 IgG (Fig. 11) fit reasonably well to the simple interaction model. Global analysis of the data gave an association rate constant of 340 M^{-1} s^{-1} and a dissociation rate constant of 3.4 x 10^{-4} s^{-1}. From these rate constants the KD of the interaction was determined to be 1.1 x 10^{-6} M.

<u>NMR studies of R3A10(Cys$^-$) and M2R2-1</u>

Both R3A10(Cys$^-$) and M2R2 were soluble up to mM concentration without precipitation or aggregation. Figure 8 shows the ^{15}N-^{1}H HSQC spectra for these two proteins. The HSQC cross peaks are well dispersed in both proton and ^{15}N dimensions, indicating the proteins are folded in solution. Excluding those from the side chain amides, ~120 HSQC cross peaks were

observed for both R3A10(Cys$^-$) and M2R2, which is less than that expected. Most of these HSQC peaks (>90%) were assigned by using heteronuclear NMR data. The cross peaks corresponding to the amides of residues in the CDR3 were found missing in the HSQC spectra, which suggests that the CDR3 in both R3A10(Cys$^-$) and M2R2-1 is either not structured or have multiple conformations in solution. By using a combination of the HNCACB and CBCA(CO)NH spectra, most of the backbone (NH, ^{15}N, $^{13}C^{\alpha}$) and side-chain $^{13}C^{\beta}$ resonances were assigned for the residues having HSQC cross peaks. The protein secondary structure was analyzed using $^{13}C^{\alpha}$ chemical shifts for the assigned residues (Wishart and Sykes, 1994). Most of $^{13}C^{\alpha}$ resonances were down-field shifted when compared with the corresponding random coil values, suggesting that the proteins are rich in β-strands. This is in agreement with the β-structures typically formed for immunoglobulin variable domains.

Minimizing the size of antigen binding proteins to a single immunoglobulin domain has been one of the primary goals of antibody engineering over the past decade. However, low levels of soluble expression in *E. coli* and solubility problems have hampered development of such molecules. The discovery of camelid heavy chain antibodies (Hamers-Casterman et al, 1993) opened up new opportunities for development of single domain antibodies, including the incorporation of features of these antibodies into human V_H frameworks. Camelization of human V_Hs is a promising technology for the generation of small antigen binding fragments that should be useful for therapeutic purposes in humans. However, while the camelized antibodies described in the literature (Davies and Reichmann, 1994; Davies and Reichmann, 1995; Reichmann and Davies, J. (Biomolecular NMR) have tremendously improved physical

properties relative to their non-camelized counterparts, these properties are still less than ideal.

Davies and Reichmann camelized a human V_H by introducing Gly44Glu, Leu45Arg and Trp47Gly mutations. However, the yields in *E. coli* of soluble camelized product were low (typically less than 1mg/l) and in order to obtain the yields and stability required for NMR studies they opted for a Trp47Ile mutation instead of the Trp47Gly mutation (Davies and Reichmann, 1995). This resulted in yields of up to 5 mg/l which is an order of magnitude lower than the yields reported here for camelized BT32/A6. A NMR structure of a human VH camelized in this manner has been described (Reichmann, J. Mol Biol) but in order to reduce aggregation and achieve sufficient solubility CHAPS detergent had to be added to the sample during NMR data collection. By contrast the camelized BT32/A6 molecules described here were completely free of aggregated material in the absence of detergent. Size exclusion chromatograms showed single peaks at an elution position expected for monomer V_H and high quality NMR data was collected in the absence of detergent.

A6VH displays a number of features that makes it a desirable template for camelized library construction. First, both its expression and its solubility are very high, atypical of VHs which are derived from conventional four chain antibodies. Second, the protein is mostly existed in a monomeric form (Figure 6B). Third, it had an unusually long CDR3 and therefore approximates the V_HH situation.

As a template for camelized V_H library construction, BT32/A6 offers the option for introduction of a CDR1-CDR3 disulphide bridge. Formation of the disulphide bond was confirmed for several sequences and introduction of the two cysteines did not have a negative

impact of the yield of soluble product. For M-2 binders the presence of the disulfide imposed a constraint that prevented optimal interaction with M-2. In other instances, however, the presence of the bridge would probably be advantageous since this is a common feature of heavy chain antibodies. Construction and pooling of two libraries, one with and one without the bridge, would appear to be advantageous.

The observation that the consensus sequence recognized by M-2 (XYKXXD) always occurred in the C-terminal half of the CDR3 of the M-2 binders is thought to indicate that contact residues reside in this portion of the CDR. On a random basis and considering the length of the randomized region of CDR3 the consensus sequence should occur at a frequency of 4×10^{-4}. Consequently, a library with 2×10^{7} members should contain 500 independent anti-M2 dAbs displaying the consensus FLAG sequence on CDR3. The preferential use of the C-terminus as an antigen contact region is in sharp contrast to an anti-lysozyme dAb where all the antigen contacting residues of CDR3 are located at its N-terminal half (Desmyter, Transue, et al., 1996),

It is not surprising that monovalent display using a phagemid vector failed to yield binders. Davies and Riechmann (1995) also constructed a camelized dAb library by randomizing CDR3 amino acid residues but the library was ten times larger and yielded anti-hapten dAbs with dissociation constants in the range of 100-400 nM. However, the isolated anti-protein dAbs had weak affinity (Davies & Riechmann, 1995) (Davies & Riechmann, 1996). Therefore, a smaller library such as the one constructed here may therefore contain only weak anti-protein dAbs. The isolation of such dAbs would be difficult with monovalent display (Lowman, Bass, et al., 1991). In a phage vector format the dAb are displayed 3-5 copies and therefore there is potential for avidity which increases the likelihood of isolating weak binders (Nissim, Hoogenboom, et al., 1994).

In other preferred embodiments of the invention, the randomized positions in A6 and A6.1 libraries are preferably at positions 100i to 100n as indicated by the data demonstrating binding in the C-terminal regionof the CDR3 loop.

In addition to CDR3 residues, CDR1 positions could be identified for limited randomization. Libraries containing shorter and partially randomized CDR3 could be constructed and pooled to further increase diversity.

Figure 15 is a schematic representation of the steps taken to remove the recombination site at the 5' end of the A6VH gene. Using the plasmid pSJF-A6VH as template and 1 & 3 and 2 & 4 primer pairs, two overlapping fragments were constructed by PCR. From these, a larger construct (Fgmt1) was assembled by splice overlap extension (SOE) and further amplified by PCR using primers 1 and 2. For simplicity only the part of the plasmid spanning from RP (1) primer binding site to FP (2) primer binding site and containing the A6VH gene is shown. 3=Chi.F primer; 4=Chi.R primer (example 21).

Additional embodiment of the randomization strategy for the libraries of the invention, are described below.

The present inventors have also found a method of enhancing the probability that the binding fragments displayed in the library have characteristics which approximate the desired solubility characteristics found in the wild type binding fragment. During construction of the library, nucleotides of the variable region are added in a step-wise addition and by selecting a

nucleotide ratio which is biased in favor of producing amino acids which reflect the DNA of the parental or wild type species.

Thus, a method for biasing a library in favor of obtaining selected percentages of wild type amino acid residues is achieved by creating residue substitutions by using different spiking levels of the various dNTPs as described below. When creating a phage library, the randomization of amino acids is often achieved by DNA synthesis. A primer is annealed next to DNA encoding for the variable region, and nucleotides are randomly added to synthesize randomized variable regions. Normally, at the step of synthesizing the DNA used to produce the variable region of the phage library, one uses a nucleotide ratio of 1:1:1:1, which generates a totally random variable region. By the present method, during synthesis of the variable region, the likelihood of achieving affinity or other desirable traits found in the wild type as follows. At each step of adding a nucleotide to the DNA variable region, one selects a dNTP ratio which is biased in favor of producing amino acids which reflect the DNA of the parental (wild type) species.

Table 3 charts particular amino acid residues or sequences of residues and preferred types of amino acid substitutions according to various examples of the invention to be defined hereafter. The selection of amino acids for randomization or partial randomization is based on adopting one or more of a variety of approaches including one of more of the following:

1. universal recognition of wild-type amino acids through a broad-based biasing in favour of the wild-type amino acids in one or more regions of interest (approximately

10%-90% biasing) in order to maintain the characteristics of the parental V_H ligand-bindingmolecule;

2. selective recognition of amino acids that are important to maintain as wild-type through biasing (approximately 10-100%) in order maintain conserved or strategic regions of amino acid residues of the parental V_H ligand-binding molecule; and

3. recognition of selected amino acids as important for intermolecular interaction and biasing those amino acids to wild-type and amino acids of the same type.

Table 3

Amino Acid Residue #s	**Description of Various Preferred Amino Acid Constitutions**
a. At least one of 100a-100h, preferably at each position of 100a-100h	Randomize; At least approximately 10% biasing in favor of wild-type amino acids; At least approximately 50% biasing in favor of wild-type amino acids; At least approximately 90% biasing in favor of wild-type amino acids; Randomize, but bias 100f to wild-type (approximately 10-100%)

b. At least one amino acid of: 100a-100b and 100g-100h preferably at each position of 100a-100b and 100g-100h	Randomize; Randomize with bias to wild-type (approximately 10-100%), preferably at least approximately 50% wild-type, alternatively at least approximately 90% wild-type amino acids;30 Randomize with bias to one of the amino acids selected from the group consisting of tyrosine, histidine, glutamine, asparagine, lysine, aspartic acid and glutamic acid (approximately 10-100%)
c. At least one of 100b-100g, preferably at each position of 100b-100g	Randomize; Delete;
d. 100a-100h	Random additions of up to 10 amino acids; Random deletions of up to 7 amino acids;
e. 95-100o	Randomize; Random additions of up to 10 amino acids; Random deletions of up to 7 amino acids;
f. At least one of 95-100, preferably at each position of 95-100	Randomize; Randomize with bias to wild-type (approximately 10-100%), preferably at least approximately 50% wild-type, or preferably at least approximately 90% wild-type amino acids;

	Invariant (primer spans this region)
g. 101-102 conserved amino acids	Invariant (primer spans this region) N/A
h. 100I-100o	Randomize Randomize with bias to wild-type (approximately 10-100%), preferably at least approximately 50% wild-type, more preferably at least approximately 90% wild-type amino acids; Randomize with bias to one of the amino acids selected from the group consisting of tyrosine, histidine, glutamine, asparagine, lysine, aspartic acid and glutamic acid (approximately 10-100%); Randomize with bias to maintaining 100o as wild-type (10-100%).
i. At least one amino acid of 100a-100b, 100g-100h and 100l-100o, preferably at each position of 100a-100b, 100g-100h and 100l-100o	Randomize with bias to wild-type (approximately 10-100%), preferably at least approximately 50% wild-type, more preferably at least approximately 90% wild-type amino acids; Randomize with bias to one of the amino acids selected from the group consisting of tyrosine, histidine, glutamine, asparagine, lysine, aspartic acid and glutamic acid (approximately 10-100%); Bias to aromatic amino acids (10-100%)
j. 95-100h	Randomize but maintain any 5-10 consecutive amino acids as wild-

	type
k. 100a-100o	Randomize but maintain any 5-10 consecutive amino acids as wild-type

Unless otherwise necessarily implied as a result of logistical considerations, it is to be understood that the various embodiments which relate to choice of amino acids for random, biased or fixed substitution (specified in column 1) as well as the various embodiments related to types of substitutions (column 2) are not mutually exclusive. Moreover the various permutations and combinations of such substitutions are hereby contemplated as embodiments of the invention. For example, substitutions referred to in row a. (any one or more amino acids and preferably all amino acids of residues 100a – 100h) #3 (at least approximately 50% wild-type amino acids) may combined with row b. (any one or more and preferably all of amino acids residues 100a, 100b, 100g and 100h) #2 (for instance, at least approximately 90% wild-type amino acids) so that, for instance, any 3 of the amino acids in 100a – 100h are biased in favor of wild-type in approximately 50% of the variant V_H ligand-binding fragments and 100a and 100b are biased in favor of wild-type in 90% of potential binding fragments.

By necessary implication the three amino acids that are biased in favor of wild-type are not residues 100a and 100b, but they may be any other three residues. Accordingly, the broadest possible interpretation is to be given to the disclosure of the various combinations and

permutations of the embodiments disclosed herein. Furthermore, it is to be understood that each of the various embodiments described herein are disclosed, except insofar as logistically impossible, in reference to each of the various aspects and definitions of the invention. Moreover, it is to be understood that phrases such as at least approximately 10%, or approximately 10-100% are intended to specify a preference for each of the unit percentages between about 7 and 100% that are practically achievable by oligonucleotide primer design and PCR amplification described herein below, as well as other well known PCR techniques and techniques of Controlled Mutation described in the art, and routine variations of such techniques. By the same token, phrases such as at least 80% are intended to specify a preference for each of the unit percentages between 80% and 100%. It is to be understood that biasing of a percentage less than 100% implies unless otherwise implied or stated that the remaining percentage is fully randomized. Furthermore, it is to be understood, for example, that 90% biasing in favor of wild-type amino acids at a given amino acid position is to be approximated by controlling the percentage amounts of each of the three relevant nucleotides (so that, for example, the product of the probabilities of occurrence of the three desired nucleotides in sequence in the growing chain is 90%) so as to supply 90% of correct coding triplet(s) and a total of 10% of random coding triplets, having regard to the degeneracy of the genetic code (for example if two different coding triplets result in a given amino acid, then the sum of the probabilities of achieving those two triplets will have to equal 90%). This is preferably accomplished on an amino acid by amino acid basis so that, for example the probability of achieving two and three wild-type amino acids in sequence, in the case of 90% biasing is 0.81 and 0.73, respectively, etc. It is to be understood that this high level of biasing may be suitable only for part of the coding sequence into which variability is introduced and

that higher levels of biasing are acceptable, when for example substantially all of the amino acids of a long CDR3 are biased, as disclosed in one of embodiments herein.

Accordingly there is a balance to be struck between a large diverse library and biasing for multifactorial characteristics such as solubility. Nevertheless it is contemplated that the library produced may be a pooled library in which several libraries each having varying degrees of biasing to wild-type, for example, 60%, 50%, 40% and 30%, are pooled together to obtain the both desired variability and similarity. The preferred parental binding-fragment may be engineered to maximize the desired characteristic (e.g. solubility, intermolecular interaction) and then made the subject of libraries with varying degrees of biasing. In this connection, the library could be biased to be rich in amino acids, which are highly soluble. It is to be understood that both arms (halves) of the preferred longer loop forming CDR3s may be biased to amino acids that are favored for intermolecular interaction, preferably charged amino acids, so as to provide a method of generating, in addition to loop size, varying loop structures. This bias may be systematically introduced or systematically reduced by randomization, in cooperating pooled libraries having varying degrees of biasing.

With respect to the application of these methods to parental V_H, preferably, CDR3s of a variety of different lengths from 16 to 33 amino acids are predominantly represented among the variant V_H ligand-binding fragments. Preferably CDR3s of a variety of different lengths, from 18 to 25 amino acids, or, from 18 to 23 amino acids are predominantly represented in the library. Although the term "predominant" ordinarily implies a majority representation of the specified long CDR3 variant V_H ligand-binding fragments, the invention also contemplates an even less substantial representation, especially within a reasonably large size

library (>10^7). Preferably, the specified long CDR3 variant V_H ligand-binding fragments have a majority representation within the library and more preferably an even greater or exclusive representation.

Optionally, the parental V_H ligand-bindingmolecule is reduced in size and the parental V_H ligand-binding molecule is optionally modified by deleting a portion of the CDR2. In another embodiment, CDR3s of the same length as that of the parental V_H ligand-binding moleculeare predominantly or exclusively represented in the variant V_H ligand-binding fragments.

In another aspect, the CDR3 region is specifically retained along with human sequence elements of other regions that confer favorable characteristics solubility, to create a phage display library having favorable characteristics of solubility, preferably when compared with variant V_H ligand-binding fragments that have fully randomized hypervariable regions (particularly CDR3). In particular, the present inventors have found that favorable solubility characteristics of a parental V_H ligand-binding molecule can be maintained in the population of variant V_H ligand-binding fragments in the course of randomizing the hypervariable regions by biasing all or selected amino acids residues to wild-type and/or biasing in favor of amino acids residues that favor certain or a variety of types of intermolecular interaction. This is respectively accomplished by increasing the percentage amounts of nucleotide bases that represent wild-type amino acids and/or amino acids that provide favorable intermolecular interactions during the randomization procedure e.g. site directed PCR mutagenesis.

Thus, variant V_H ligand-binding fragments having relatively long CDR3s of varying lengths are produced by randomly or partially randomly inserting varying numbers of nucleotide triplets in any part of a randomized portion of the parental V_H framework. Primers of the desired length and nucleotide composition are synthesized followed by PCR amplification. Desired randomization can be achieved by biasing nucleotide composition of the primer. The production of displays of long CDR3 variant binders may also be accomplished by pooling several libraries of variantV_H ligand-binding fragments having randomized or partially randomized CDR3s of different respective uniform lengths. These strategies are not mutually exclusive.

The additional following terms are used herein as follows, unless the context logically implies otherwise:

"Biasing", "biased in favor of" and related forms of these terms are generally intended to refer to weighting in the course of introducing variation in the parental ligand-binding molecule.

"Homologous" or "homology" as used herein refers to "identity" or "similarity" as used in the art, meaning relationships between two or more polynucleotide or amino acid sequences, as determined by comparing the sequences. In the art, identity also means the degree of sequence relatedness between polynucleotide sequences, as the case may be, as determined by the match between strings of such sequences. Both identity and similarity can be readily calculated (Lesk, A. M., ed., *Computational Molecular Biology*, Oxford University Press, New York, 1988; Smith, D. W., ed., *Biocomputing: Informatics and Genome Projects*,

Academic Press, New York, 1993; Griffin, A. M., and Griffin, H. G., eds., *Computer Analysis of Sequence Data, Part I*, Humana Press, New Jersey, 1994; von Heinje, G., *Sequence Analysis in Molecular Biology*, Academic Press, 1987; and Gribskov, M. and Devereux, J., eds., *Sequence Analysis Primer*, M Stockton Press, New York, 1991). While there exist a number of methods to measure identity and similarity between two polynucleotide sequences, both terms are well known to skilled artisans (von Heinje, G., 1987;Gribskov, M. and Devereux, J., 1991; and Carillo, H., and Lipman, D., 1988). Methods commonly employed to determine identity or similarity between sequences include, but are not limited to those disclosed in Carillo, H., and Lipman, D. (1988, *SIAM J. Applied Math.*, 48: 1073). Methods to determine identity and similarity are codified in computer programs. Preferred computer program methods to determine identity and similarity between two sequences include, but are not limited to, GCG program package (Devereux, J., et al.(1984), *Nucleic Acids Research* 12(1): 387), BLASTP, BLASTN, and FASTA (Atschul, S. F. et al.(1990), *J. Molec. Biol.* 215: 403). "Percent homology" or "% homologous" or related terms include both of the following interpretations / methods of calculation: 1) an approximate percentage of the sequence referenced in terms of the number of common residues (e.g. 80% of 11 is understood to be an approximation insofar as application of the percentage does not yield a unit number of residues, in which case both the immediately higher number and immediately lower unit numbers, 9 and 8 respectively, are deemed to be covered); 2) the percentage of binding fragments theoretically achievable that have the full wild-type sequence, which is calculated as a product of the probabilities that the wild-type amino acid will occur at a given amino acid position.

"Conserved" regions refer to those which are commonly found in at least other antibodies of the same type or in at least the same species of mammal.

"Wild-type" refers to the parental binding-fragment, which may be a variant of the natural or to the native A6 V_H parental ligand-binding fragment, depending on the context.

"Step-wise" refers to the addition of, for example, nucleic acids, in a manner such that the quantity of nucleic acids added at each step is rigorously control, usually one nucleic acid at a time.

"Spanning" does not preclude deletions or additions within the parental V_H binding-fragment that are not inimical to the operation of the invention.

"Camelid type" refers specifically to one or more features of the camelid V_L interface.

"Soluble" includes the generally ascribed meaning in the art and without limitation includes (based on solubility correlated phenomena) the relative amounts of naturally-folded recombinant protein released from the cell.

"Percent biasing" or "% of binding fragments" (or "biasing 10-100%", etc.) refers to biasing on an individual amino acid basis (though other techniques to accomplish the same effect might apparent to those skilled in the art). Similarly, the specification that wild-type amino acids occur at a specified position or series of positions in, for example, at least approximately 50% of potential binding fragments is intended to mean both that 50% biasing

is sought at a given such position or that a total of 50% of the correct nucleotide triplets are represented.

"Approximately" in reference to percentages is intended to accommodate attrition of various desired variant V_H ligand-binding fragments, the assumption that the probabilistic outcomes will not be achieved in practice and that certain variation in methods to accomplish the specified results is deemed to be suitable. The term 50% in reference to an uneven number of amino acids residues means that either one more or one less than half of the amino acids is referred to.

The practice of the present invention employs, unless otherwise indicated, conventional techniques of molecular biology (including recombinant techniques), microbiology, cell biology, biochemistry and immunology, which are within the skill of the art. Such techniques are explained fully in the literature, such as, "Molecular Cloning: A Laboratory Manual", second edition (Sambrook et al., 1989); "Oligonucleotide Synthesis" (M.J. Gait, ed., 1984); "Animal Cell Culture (R.I. Freshney, ed., 1987); "Methods in Enzymology" (Academic Press, Inc.); "Handbook of Experimental Immunology" (D.M. Wei & C.C. Blackwell, eds.); "Gene Transfer Vectors for Mammalian Cells" (J.M. Miller & M.P. Calos, eds., 1987); "Current Protocols in Molecular Biology (F.M. Ausubel et al., eds., 1987); "PCR: The Polymerase Chain Reaction", (Mullis et al., eds., 1994); "Current Protocols in Immunology" (J.E. Coligan et al., eds., 1991). These references are incorporated herein by reference. These techniques are applicable to the production of the polynucleotides and polypeptides of the invention, and, as such, may be considered in making and practicing the

invention. Particularly useful techniques for particular embodiments will be discussed in the sections that follow.

Recombinant genetic techniques have allowed cloning and expression of antibodies, functional fragments thereof and the antigens recognized. These engineered antibodies provide novel methods of production and treatment modalities. For instance, functional immunoglobulin fragments have been expressed in bacteria and transgenic tobacco seeds and plants. Skerra (1993) *Curr.Opin. Immunol.* 5:256:262; Fiedler and Conrad (1995) *Bio/Technology* 13:1090-1093; Zhang et al. (1993) *Cancer Res.* 55:3384-3591; Ma et al. (1995) *Science* 268:916; and, for a review of synthetic antibodies, see Barbas (1995) *Nature Med.* 1:836-839. These and more current references describing these techniques, which these references, particularly those well known to persons practicing in the relevant arts, are hereby incorporated herein by reference.

Suitable parental binding-fragments include any known in the art and include the group consisting of an scFv, Fab, V_H, Fd, Fabc, $F(ab')_2$, $F(ab)_2$ derived from A6.

Nucleotide sequences can be isolated, amplified, and processed by standard recombinant techniques. Standard technique in the art include digestion with restriction nucleases, and amplification by polymerase chain reaction (PCR), or a suitable combination thereof. PCR technology is described in U.S. Patent Nos. 4,683,195; 4,800,159; 4,754,065; and 4,683,202, as well as *PCR: The Polymerase Chain Reaction*, Mullis et al., eds., Birkauswer Press, Boston (1994).

In addition to the specific PCR methods of biasing to wild-type A6 amino acid residues detailed below, it is possible to produce multiple different oligonucleotide primers consisting of specified amino acid residues (one or more) of the wild-type A6 molecule (e.g. CDR3 residues), mixing these in appropriate concentrations with a completely randomized (e.g. CDR3) oligonucleotide primer and subjecting the mixture of oligonucleotide primers to PCR. This will result in a biased phage library population of one's choosing (i.e. the amounts of the selectively randomized and totally randomized primers in the mixture will determine the per cent of each CDR3 representation in the library).

Polynucleotides comprising a desired sequence can be inserted into a suitable vector, and the vector in turn can be introduced into a suitable host cell for replication and amplification. Polynucleotides can be introduced into host cells by any means known in the art. Cells are transformed by introducing an exogenous polynucleotide by direct uptake, endocytosis, transfection, f-mating or electroporation. Once introduced, the exogenous polynucleotide can be maintained within the cell as a non-integrated vector (such as a plasmid) or integrated into the host cell by standard methods. See, e.g., Sambrook et al. (1989). RNA can also be obtained from transformed host cell, or it can be obtained directly from the DNA by using a DNA-dependent RNA polymerase.

Suitable cloning and expression vectors include any known in the art, e.g., those for use in bacterial, mammalian, yeast and insect expression systems. Specific vectors and suitable host cells are known in the art and are not described in detail herein. See e.g. Gacesa and Ramji, *Vectors*, John Wiley & Sons (1994).

Phage display techniques are generally described or referenced in some of the preceding general references, as well as in U.S. Patent Nos. 4,593,002; 5,403,484; 5,837,500; 5,571,698; 5,750,373; 5,821,047; 5,223,409 and 5,702,892. "Phage Display of Peptides and Proteins", (Kay, Brian K. et al., 1996); "Methods in Enzymology", Vol. 267 (Abelson, John N., 1996); "Immunology Methods Manual", (Lefkovits, Ivan, 1997); "Antibody phage display technology and its applications", (Hoogenboom, Hennie R. et al., 1998). *Immunotechnology* 4 p.1-20; Cesareni G et al. Phage displayed peptide libraries. *Comb Chem High Throughput Screen*. 1999 Feb;2(1):1-17; Yip, YL et al. Epitope discovery using monoclonal antibodies and phage peptide libraries. *Comb Chem High Throughput Screen*. 1999 Jun;2(3):125-38; Rodi DJ et al. Phage-display technology--finding a needle in a vast molecular haystack. *Curr Opin Biotechnol*. 1999 Feb;10(1):87-93.

Generally, DNA encoding millions of variants of a parental binding-fragment can be batch-cloned into the phage genome as a fusion to the gene encoding one of the phage coat proteins (pIII, pVI or pVIII). Upon expression, the coat protein fusion will be incorporated into new phage particles that are assembled in the bacterium. Expression of the fusion product and its subsequent incorporation into the mature phage coat results in the ligand being presented on the phage surface, while its genetic material resides within the phage particle. This connection between ligand genotype and phenotype allows the enrichment of specific phage, e.g. using selection on immobilized target. Phage that display a relevant ligand will be retained, while non-adherent phage will be washed away. Bound phage can be recovered from the surface, reinfected into bacteria and re-grown for further enrichment, and eventually for analysis of binding. The success of ligand phage display hinges on the combination of

this display and enrichment method, with the synthesis of large combinatorial repertoires on phage.

While the use of phage is described as an embodiment for the production of libraries for displaying, and selecting particular binding fragments, it is to be understood that and suitable genetic package may be used for the production of libraries of the invention. Such suitable genetic packages include cells, spores and viruses (see US Patent No. 5,571,698), or any other suitable replicable genetic packages. With respect to cell based approaches, another popular method of presenting a library is the two-hybrid system (Feilds and Sternglanz, 1994, *Trends in Genetics* 10:286-292). Those skilled in the art will appreciate that in vitro systems (non-cell based) may be equally applicable to the methods of the present invention, for example ribosome display (Hanes et al., 1998) or RNA-peptide fusion (Mattheakis et al., 1994, Proc Natl Acad Sci USA 91:9022-26; Hanes et al., 1999, Curr Top Microbiol Immunol 243:107-22).

Ribosome display is a well documented technique that may be useful for generating libraries. This entirely in vitro method allows for libraries with a diversity of $>10^{12}$. In this method, a peptide is displayed on the surface of a ribosome that is translating it. Briefly, a library of mRNA molecules (we could start with A6) is translated in vitro translation system to the 3' end, such that the ribosome does not fall off. The protein emerges from the ribosome in such a way that it can fold, but does not fall off. In some instances, there is an additional folding step in an oxiding environment (important for proteins with disulfide bonds). The whole complex of folded protein, ribosome and mRNA, which is stable for several days, can then be panned against a ligand that is recognized by the translated protein. (For example, the

translated protein could be an antibody and the ligand is its antigen). The mRNA can then be amplified by reverse transcription and PCR. This technique has been used to successfully generate scFv antibody fragments with high affinity for their target. Reference is made to Hanes,J., Jermutus,L., Weber-Bornhauser,S., Bosshard,H.R. & Pluckthun,A. Ribosome display efficiently selects and evolves high-affinity antibodies in vitro from immune libraries. *Proc. Natl. Acad. Sci. USA* **95**, 14130-14135 (1998); Schaffitzel,C., Hanes,J., Jermutus,L. & Pluckthun,A. Ribosome display: an in vitro method for selection and evolution of antibodies from libraries. *Journal of Immunological Methods* **231**, 119-135 (1999); He,M. *et al.*. Selection of a human anti-progesterone antibody fragment from a transgenic mouse library by ARM ribosome display. *Journal of Immunological Methods* **231**, 105-117 (1999); Roberts,R.W. Totally in vitro protein selection using mRNA-protein fusions and ribosome display. *Current Opinion in Chemical Biology* **3**, 268-273 (1999); Williams,C. Biotechnology match making: screening orphan ligands and receptors. *Current Opinion in Biotechnology* **11**, 42-46 (2000); Mattheakis,L.C., Bhatt,R.R. & Dower,W.J. An in vitro polysome display system for identifying ligands from very large peptide libraries. *Proc. Natl. Acad. Sci. USA* **91**, 9022-9026 (1994).

Example 1 - Construction of Single-domain A6-based (A6-based dAb) DNA Templates

To facilitate construction of the A6-based dAb libraries, a *Nhe*I site was introduced at the amino acid residues 24-25 (nucleotides underlined and bolded in Figure 14) while maintaining the wild-type amino acid sequence. Briefly, the A6 V_H gene was used as a PCR template to amplify a shorter internal fragment employing the primers A6V_H/NheI-5'(TGTTCAGCTAGCGGATTC)3' and A6V_H/BstEII-

5'(TGAGGAGACGGTGACCGTTGTCCCTTGGCCCCAGATATCAAA)3'. These primers incorporate *Nhe*I and *BstE*II sites (underlined) at the 5' and 3' ends of the amplified product. PCR (polymerase chain reaction) was performed in a total volume of 50 :l containing 200 mM each of the four dNTPs, 100 pmol each of the two primers, 5 l of 10X buffer (New England Biolabs (NEB)), and 2 units of Vent DNA polymerase (NEB).

The amplified product was purified using QIAquick PCR Purification kit™ (QIAGEN, Mississauga, ON), digested with *Nhe*I and *BstE*II restriction endonucleases and subsequently ligated to the *Nhe*I/*BstE*II-restricted pSJF1-10A12 vector derived from pUC 8 (Narang et al., 1987) to replace a portion of the existing A6 V_H gene. To construct the pSJF1 vector, the pUC 8 plasmid (Vierra and Messing, 1982; Messing, 1983) was modified by inserting the OmpA signal sequence and the His_5-carboxy tail between the EcoRI and HindIII restriction sites of the pUC 8 polylinker region, using oligonucleotide primers and PCR (Narang et al., 1987).

Electro-competent *E.coli* TG1 cells were prepared (Tung and Chow, 1995) and an aliquot of the ligated product was used to transform the cells. Transformation was carried out using the BIO-RAD Gene Pulser™ (Bio-Rad Laboratories, Mississauga), ON according to the manufacturer's instructions and the clone harbouring the mutated A6 dAb gene was confirmed by sequencing (Sanger, F. et al., 1977) using the AmpliTaq DNA Polymerase FS kit and 373A DNA Sequencer Stretch (PE Applied Biosystems, Mississauga, ON). All the cloning steps were performed as previously described (Sambrook et al. 1989). The resulting vector is termed pSJF1-A6VH.*Nhe*I.

Example 2 - A6 dAb Library Construction

The steps of A6 dAb library construction involved a series of sequential PCR experiments.

(1) Introduction of restriction sites to facilitate cloning: To amplify the target DNA, the PCR mixture was first incubated at 95°C for 5 min, then subjected to 30 cycles of: 30 sec at 94°C, 1 min at 40°C and, 1 min at 72°C. The A6VH.*Nhe*I-containing plasmid, pSJF1-A6VH.*Nhe*I, was used as the template in PCR to amplify a shorter fragment using the primers A6VH/*Apa*lI – 5' (CATGACCACAGTGCACAGGAGGTCCAGC-TGCAGGAGTC) 3' and A6VH.FR3.F - 5' (TTTCACACAGTAATACAC) 3'. The PCR mixture contained 200 M each of the four dNTPs, 0.2 pmol/ 1 each of the two primers, 1X buffer (Perkin Elmer), and 0.05 units/ 1 of AmpliTaq DNA polymerase (Perkin Elmer). (The former primer also introduces *Apa*lI site at the 5' end of the PCR product.)

Example 3 - Randomization of the A6dAb CDR3 residues:

The amplified fragment from step (1) was purified by QIAquick Gel Extraction kit™ (QIAGEN) and subsequently used as the template in a second PCR reaction using the primers A6VH.*Apa*lI and A6VH.RndmCDR3.F - 5' (GCCCCAGATATCAAA20 [((A/C)NN)]TTTCACACAGTAATA)3'. At the protein level the second primer results in the randomization of the first 20 residues in CDR3. The PCR mixture was identical to above except that the concentration of the primers was increased to 0.5 pmol/l to ensure that sufficient amounts of oligonucleotide primers and dNTPs were provided for the generation of a large randomized library.

Example 4 - Addition of a Not1 restriction site, ligation to the phage vector and library construction.

The amplified fragments were purified as above and used as templates in a third round of PCR employing 2 pmol/ul each of the two primers A6VH/*ApaI* (described above) and A6VH.NotI.EXT.F – 5'(CGATTCTGCGGCCGCTGAGGAGACGGTGACCGTT-GTCCCTTGGCCCCAGATATCAAA) 3'. (The latter primer incorporates the *Not*I site (underlined) at the 3' end of the amplified products.) The amplified fragments were purified using QIAquick PCR Purification kit™ (QIAGEN), digested with *ApaI* and *Not*I, and ligated to *ApaI*/*Not*I-digested fd-tet phage vector (McAfferty et al., 1990; Zacher et al., 1980). The ligated product was desalted using QIAquick PCR Purification kit™ (QIAGEN). To determine the size of the library, immediately following the transformation and after the addition of the SOC medium (per L: bacto-tryptone, 20 g; bacto-yeast extract, 5 g; NaCl, 0.5 g; glucose, 3.6 g) a small aliquot of the electroporated cells were serially diluted in exponentially growing E. *coli* strain TG1 cells. Two hundred µl of the diluted cells were mixed with 3ml of 50°C top agar and immediately poured onto 2xYT (per L: bacto-tryptone, 16 g; bacto-yeast extract, 10 g; NaCl, 5 g) plates pre-warmed to 37°C. Plates were incubated overnight at 37°C and the number of plaques were used to determine the size of the library. Following this, the DNA inserts from single plaques were amplified using PCR. The size of the amplified product, determined by agarose gel electrophoresis, was used to determine the fraction of the library with full-sized A6 dAb inserts. Diversity of the library was determined to be in the range of 10^7-10^9.

The recombinant phage vectors, 1.5 μg, were mixed with 40 μl of competent *E. coli* strain TG1 and the cells were transformed by electroporation. Following transformation, 1 ml of SOC medium was added to each electroporation mixture (45 ml in total). The mixture was divided into three equal aliquots, each of which were added to tubes containing 3 ml of top agar at 50°C, vortexed immediately, poured onto pre-warmed 2xYT agar plates, and incubated at 37°C overnight. Five ml of sterile PBS (per L: NaCl, 8 g; KCl, 0.2 g; Na_2HPO_4, 1.44 g; KH_2PO_4, 0.24 g; pH 7.4) was added to the plates and the phage particles were eluted by gently shaking the plates at 4°C for 3 hr. The phage-containing PBS supernatant was collected, the plates rinsed with an additional 5 ml of PBS and the two supernatants were pooled. The supernatants were centrifuged at 6000g for 15 min at 4°C, the cleared supernatant decanted and the phage were purified as described by Harrison et al. (1996). The phage pellet was dissolved in 20 ml of sterile PBS, divided into 100 μl aliquots and stored in liquid nitrogen.

Example 5 - Partial Construction of A6.1 analogue

A6 dAb was constructed from the heavy chain variable domain (VH) of the A6, an anti-tumor IgM with unidentified antigen (Dan et al., 1995). Briefly, the dAb gene was amplified by polymerase chain reaction (PCR) using the primers:
H11MB, 5'(TATGGATCCTGAGGAGACGGTGACCGT)3'; and
A6VH. , 5'(TATGAAGACACCAGGCCGAGGTCCAGCTGCAGGAG)3' which contain the *Bam*HI and *Bbs*I sites (underlined). PCR was performed in a total volume of 50 l containing 200 M each of the four dNTPs, 100 pmol each of the two primers, 5 l 10X buffer (NEB), and 2 units of Vent DNA polymerase (NEB). The amplified product was

purified using QIAquick PCR Purification kit™ (QIAGEN), digested with *Bam*HI and *Bbs*I restriction endonucleases, and subsequently ligated to the expression vector pSJF2 (Simon J. Foote, personal communication). Transformation was performed as described in Example 1. To modify A6 dAb, the vector containing A6 dAb gene (pSJF2-A6dAb) was used as template to amplify two overlapping 5' and 3' fragments. The 5' fragment was amplified using the RP primer 5'(GCGGATAACAATTTCACACAGGAA)3' and A6.1dAb analogue .bk primer 5'(AGCCTGGCGGACCCAGTGCATAGCATAGCTACTGAAGGTGAATCCGCTAGCTG AACAGGAGAGTCT)3'. The 3' fragment was amplified using the FP primer 5'(CCAGGGTTTTCCCAGTCACGAC)3' and the mutagenic primer A6.1 dAb analogue fw, 5'(TGGGTCCGCCAGGCTCCAGGGAAG**GAACGT**GAA**GGT**GTTTCAGCTATTAGT)3'. (At the protein level the b o l d codons in the mutagenic primer introduce Glu, Arg, and Gly at positions 44, 45, and 47, respectively). The two fragments were gel purified using the QIAquick Gel Extraction kit™ (QIAGEN), and a larger construct was assembled from the 5' and 3' fragments by performing splice overlap extension (Clackson, et al., 1991). Briefly, the reaction vial containing both 5' and 3' fragments, 200 M each of the four dNTPs, 5 l 10X buffer (NEB), and 2 units of Vent DNA polymerase (NEB) was subjected to 7 cycles of 1 min at 94°C and 2.5 min at 72°C. To amplify the assembled construct, RP and FP primers were added to a final concentration of 10 pmol/ l and the mixture was subjected to 25 cycles of 1 min at 94°C, 1 min at 55°C, and 1 min at 72°C. The amplified product was gel purified, digested with *Eco*RI and *Hind*III, purified again, and ligated to the *Eco*RI/*Hind*III restricted pSJF2-A6dAb. An aliquot of the ligated product was used to transform the TG1 cells and the clone harboring the A6.1 dAb was identified by sequencing. As expected, at the protein level the A6.1 dAb had acquired the following three mutations: Gly44Glu, Leu45Arg, and Tyr47Gly (Davies and Riechmann, 1994)

Example 6 - Construction of the A6.1 dAb library using phagemid vector.

As the first step, the camelized dAb gene was used as the template in PCR to amplify a shorter fragment. PCR was performed as described above using the two mutagenic primers A6VH.33C,

5'(TGTTCAGCTAGCGGATTCACCTTCAGTAGCTAT**TGT**ATGCACTGGGTCCGC)3'

containing the *Nhe*I site (underlined) and A6VH.A,

5'(**TGCTGC**ACAGTAATACACAGCCGT)3'. (At the protein level the bold codons in the mutagenic primers introduce cysteine and two alanine residues at positions 33, 93, and 94, respectively). The mutated fragment was used as the template in a second PCR using the primers A6VH.33C and A6VH.100eC,

5'(GCCCCAGATATCAAA[(A/C)NN]$_9$GCA[(A/C)NN]$_{10}$TGCTGCACAGTAATA)3'. The second primer results in the randomization of 19 residues in CDR3 and introduces a cysteine at position 100e. The amplified fragments were used as the templates in a third round of PCR employing the primers A6VH.33C and A6VH.BstEII,

5'(TGAGGAGACGGTGACCGTTGTCCCTTGGCCCCAGATATCAAA)3'. (The latter primer incorporates the *BstE*II site (underlined) at the 3' end of the amplified products.) PCR was performed as above using 0.5 pmol of template and 100 pmol of each of the two primers. The amplified fragments were purified, digested with *Nhe*I and *BstE*II, and ligated to *Nhe*I/*BstE*II-treated pSJF6-A6dAb phagemid (Simon J. Foote, personal comunication). The ligated product was desalted using QIAquick PCR Purification kit™ (QIAGEN) and used to transform *E. coli* strain XL1-Blue. Various dilutions of the transformation mixture was spread on LB/ampicillin plates and incubated overnight. In the morning, the number of

ampicillin resistant colonies were used to calculate the size of the library. Following this, single colonies were suspended in PCR mixture, and the DNA inserts were amplified. The size of the amplified product, determined by agarose gel electrophoresis, was used to determine the fraction of the library with full-size dAb insert. Diversity of the library was determined by sequencing 20 dAb genes from the library. Growth of the library was performed as described (Harrison et al., 1996).

Example 7 -Subcloning the library in the pahge vectorector

As the initial step of sub-cloning, 180 pmol of the library phagemid DNA template and 100 pmol of each of the two primers A6VH.ApaII, 5'(CATGACCACAGTGCACAGGAGGTCCAGCTGCAGGAGTC)3' and A6VH.NotI 5'(CGATTCTGCGGCCGCTGAGGAGACGGTGACCGTTG)3' were used in PCR to amplify the dAb genes. The primers are complimentary to the 5' and 3' ends of the dAb genes and incorporate *ApaI* and *Not*I restriction sites (underlined sequences) at the end of the amplified genes. The amplified products were purified, cut sequentially with *ApaI* and *Not*I restriction endonucleases, purified again, and ligated to the *ApaI*/*Not*I-treated fd-tet phage vector. Following this, 1.5 g of the desalted ligated product was mixed with 40 l of competent *E. coli* strain TG1 and the cells were transformed by electroporation. Transformation, library phage amplification and purification and library size determination were performed as in Example 4.

Example 8 - Library size determination.

To determine the size of the library, immediately following the transformation and after the addition of the SOC medium an small aliquot of the electroporated cells were serially diluted in exponentially growing TG1 cells. Two hundred l of the diluted cells was mixed with 3 ml of 50 C agarose top and immediately poured onto 2xYT plates pre-warmed to 37 C. Plates were incubated overnight at 37°C and the number of plaques were used to determine the size of the library.

Example 9 - Panning

Panning was performed using the Nunc-Immuno MaxiSorp™ 8-well strips (Nunc). Briefly, the wells were coated overnight by adding 150 l of 100 g/ m l antigen in PBS. In the morning, they were rinsed three times with PBS and subsequently blocked with 400 l P B S-2% (w/v) skim milk (2% MPBS) at 37°C for 2 hr. The wells were rinsed as above and 10^{12} transducing units phage in 2% MPBS were added. The mixture was incubated at room temperature for 1.5 hr after which the unbound phage in the supernatant was removed. The wells were rinsed 10 times with PBS-0.1% (v/v) Tween 20 and then 10 times with PBS to remove the detergent. The bound phage was eluted by adding freshly prepared 200 l 1 0 0 mM triethylamine, pipetting the content of the well up and down several times and incubating the mixture at room temperature for 10 min. The eluted phage was transfered to a tube containing 100 l 1 M Tris-HCl, pH 7.4 and vortexed to neutralize triethylamine. Following this, 10 ml exponentially growing TG1 culture was infected with 150 l e luted phage by incubating the mixture at 37°C for 30 min. Serial dilutions of the infected cells were used to determine the titer of the eluted phage as described in the previous section. The remaining of the infected cells were spun down and then resuspend in 900 l 2 x YT. The cells were mixed

in 300 l aliquots with 3 ml agarose top and the phage propagated on the plates overnight at 37°C. In the morning the phage was purified, the titer was determined, and a total of 10^{11} transducing units phage were used for further rounds of selectio

Example 10 - Expression and Purification.

Thirty ml of LB containing 100 ug/ml ampicillin was inoculated with a single colony harboring pSJF2-dAb and the culture was shaken at 240 rpm at 37°C overnight. In the morning the entire overnight culture was used to inoculate 1 liter of M9 medium supplemented with 5 g/ml vitamin B1, 0.4% casamino acid and 100 g/ml ampicillin. The culture was shaken at room temperature for 30 hr at 160 rpm and subsequently supplemented with 100 ml of 10x induction medium and 100 ul of 1M isopropylthio--D-gal a ctoside. The culture was shaken for another 60 hr, the periplasmic fraction was extracted by osmotic shock method (Anand et al., 1991), and the presence of dAb in the extract was detected by Western blotting (MacKenzie 1994). The periplasmic fraction was dialyzed extensively in 10 mM HEPES (N-[2-hydroxyethyl]piperazine-N - [2-ethanesulfonic acid]) buffer pH 7.0, 500 mM NaCl. The presence of the dAb C-terminal His_5 tag allowed a one step protein purification by immobilized metal affinity chromatography using HiTrap Chelating™ column (Phamacia). The 5-ml column was charged with Ni^{2+} by applying 30 ml of a 5 mg/ml $NiCl_2.6H_2O$ solution and subsequently washed with 15 ml deionized water. Purification was carried out as described (MacKenzie, 1994) except that the starting buffer was 10 mM HEPES buffer, 10 mM imidazole, 500 mM NaCl, pH 7.0, and the bound protein was eluted with a 10-500 mM imidazole gradient. The purityof the protein was determined by SDS-PAGE (Laemmli). To detect the presence of dimer/multimer dAb in the protein preparation,

gel filtration chromatography was performed using Superdex75 (Pharmacia) as described (Deng et al., 1995).

Example 11 - Alkylation reactions

Alkylation reactions were performed using iodoacetic acid. Briefly, 5x vol. cold acetone were added to 200 g of dAb solution and the contents were mixed, followed by centrifugation in a microfuge at maximum speed at 4°C for 10 min. The pellet was dissolved in 500 l of 6 M guanidinium hydrochloride and 55 l of 1 M Tris buffer, pH 8.0 were added. Subsequently, a 25 molar excess of DTT, relative to Cys residues, was added and the mixture was incubated at room temperature for 30 min. To this, a 2.2 molar excess, relative to DTT, of freshly-made iodoacetic was added and the reaction was incubated as described above. At the end of incubation, the alkylated product was concentrated and dissolved in 50 l of distilled water using Ultrafree-MC 10,000 NMWL filter unit according to the manufacturer's instructions (Millipore, Nepean, ON, Canada). Control experiments were identical except that DTT was replaced with water. The MWs of the iodoacetic acid-treated dAbs were determined by mass spectroscopy.

Example 12 - Surface Plasmon Resonance

Binding studies were performed using BIACORE Upgrade (Biacore Inc., Piscataway, NJ) as described (Jönsson et al., 1991). Approximately 14, 000 RU of anti-FLAG M2 IgG or control IgG were immobilized on CM5 sensor chips by amine coupling. Single-domain antibodies were passed over the sensor chips surfaces in 10 mM HEPES buffer, pH 7.4, 150

mM NaCl, 3.4 mM EDTA, 0.005% P-20 (Biacore Inc.) at 25°C and at a flow rate of 5 µl/min. To assess the effect of DTT on the dAb binding to M2, dAbs were incubated with DTT prior to injection and the above buffer was supplemented with appropriate amount of DTT. Surfaces were regenerated with 10 mM HCl. Sensorgram data were analyzed using the BIAevaluation 3.0 software package (Biacore Inc.).

Example 13 - Enzyme-Linked Immunosorbent Assay (ELISA)

Nunc-Immuno MaxiSorp™ plates (Nunc) were coated overnight at 4°C with 150 l o f 10 g/ml of 3B1 scFv or BSA in PBS. The contents were removed and the plates were tapped on a paper towel to remove any liquid remaining in the wells. The wells were blocked by adding 300 µl of 2% MPBS and incubating for 2 hr at 37°C. The contents of the wells were emptied as before, 100 l of purified dAb phage in 2% MPBS was added, and the wells were incubated at room temperature for 1.5 hr. The contents were emptied again and the wells were washed 5 times with PBS-0.05% (v/v) Tween 20 and subsequently blotted on a paper towel to remove any remaining wash buffer. One Hundred l of recommended dilution of HRP/Anti-M13 monoclonal antibody conjugate (Amersham Pharmacia Biotech) in 2% MPBS was added and the wells were incubated at room temperature for 1 hr. The wells were washed six times as before and the binding of dAb to the antigen was detected colorimetrically by adding 100 l of equal mixtures of TMB Peroxidase Substrate and H_2O_2 (Kirkegaard and Perry Laboratories, Gaithersberg, MD, USA) at room temperature for several minutes. The reaction was stopped by adding 100 l of 1 M H_3PO_4 and the A_{450} was measured by DYNATECH MR5000 plate reader (Dynatech Laboratories, Chantilly, VA, USA).

Example 14 - NMR studies

Sample preparation

Isotopically labeled proteins were prepared from cells grown on ^{15}N- and/or ^{13}C-enriched media (Bio-Express, Cambridge Isotopes Laboratory, Andover, MA). Briefly, six ml of LB containing100 ug/ml ampicilin was inoculated with a single recombinant colony and incubated at 37°C and 260 rpm to an A_{600} of about 5. The cells were centrifuged and then re-suspend in 3 ml sterile PBS and the A600 was measured. The cells were added to twenty-five ml of Bioexpress/100 ug/ml ampicilin in sterile 125 ml Erlenmyer flasks to a final concentration of A600=0.06 and Incubated at 37°C at 200 rpm for 9-10 hours. The periplasmic fraction was extracted by osmotic shock method (Anand, Dubuc, et al.. 1991) and the presence of dAb in the extract was detected by Western blotting (MacKenzie, Sharma, et al.. 1994). The periplasmic fraction was dialyzed extensively in 10 mM HEPES buffer, pH 7.0, 500 mM NaCl. The presence of a C-terminal His_5 tag allowed a one step protein purification by IMAC using HiTrap Chelating™ column (Amersham Phamacia Biotech, Baie d'Urfé, QC, Canada). The 5-ml column was charged with Ni^{2+} by applying 30 ml of a 5 mg/ml $NiCl_2.6H_2O$ solution and subsequently washed with 15 ml deionized water. Purification was carried out as described previously (MacKenzie, Sharma, et al.. 1994) except that the starting buffer was 10 mM HEPES buffer, 10 mM imidazole, 500 mM NaCl, pH 7.0, and the bound protein was eluted with a 10-500 mM imidazole gradient. The purity of the protein was determined by SDS-PAGE (Laemmli 1970). NMR samples were prepared by concentration and extensive buffer exchanging on a YM10 membrane (Amicon). The final

buffer contained 10 mM sodium phosphate, 150 mM NaCl, and 0.2 mM EDTA at pH 6.8. The final protein concentration of the NMR samples was ~ 1 mM.

NMR spectroscopy

NMR experiments were performed at 298 K and 308 K on a Bruker Avance800 spectrometer equipped with pulse field gradient accessories. 2D ^{15}N-^{1}H HSQC (Bodenhausen and Ruben, 1980) was acquired using solvent suppression via the WATERGATE method implemented through the 3-9-19 pulse train (Piotto et al., 1992; Sklenar et al., 1993). Triple-resonance experiments (Slatter et al.. 1999, and references therein) including HNCACB, CBCA(CO)NH, HNCA, HN(CO)CA, HNCO, HBHA(CO)NH, and ^{15}N-edited 3D NOESY-HSQC, 3D TOCSY-HSQC were acquired at 308 K and 298 K for R3A10 (cys-) and M2R2, respectively. The NMR data were processed using NMRPipe/NMRDraw (Delaglio et al.., 1995) and analyzed by the use of NMRView (Johnson and Blevins, 1994). Chemical shifts were referenced internally to sodium 2,2-dimethyl-2-silapentane-5-sulphonate (DSS) for proton and calculated for ^{15}N and ^{13}C assuming $\gamma^{15}N/\gamma^{1}H = 0.101329118$ and $\gamma^{13}C/\gamma^{1}H = 0.251449530$ (Wishart et al.., 1995).

Example 15 - Testing of the Phage Display A6 dAb Library Against the Anti-FLAG M2

Monoclonal IgG Antibody

The phage display A6-based dAb library was panned against the anti-FLAG M2 monoclonal antibody as described by the New England Biolabs (Beverly, MA) (NEB Technical Bulletin

(1998): Ph.D.Ô Phage Display Peptide Library Kits; Knappik and Pluckthun, 1994); the FLAG peptide epitope recognized by the M2 monoclonal antibody is (X)YKXXD where the first position has a preference for aspartic acid (Miceli et al., 1994). On a random basis, considering the length of the randomized region of A6 CDR3 (i.e., 20 residues), the consensus sequence should occur at a frequency of $4x10^{-4}$. Thus, in the A6 dAb library with $2x10^{7}$ individual clones, the FLAG peptide epitope should be represented by approximately $5x10^{2}$ independent clones.

After three rounds of panning against M2 IgG thirty one clones from rounds two and three were selected and their A6 dAb genes sequenced. Twelve different A6 dAb genes with the FLAG consensus sequence were identified (Table 1, first twelve entries).

Example 16 - Introducing Genetic Variation into the Sequence Corresponding to the A6 Heavy Chain CDR3 Region - Randomized Residues

Oligonucleotides comprising randomly mutated CDR3 regions were prepared on an Applied Biosystems 394 DNA synthesizer as described above.

1. Production of 23 randomized residues (CDR3 1-23):

The anti-codon formula [(A/C)NN] is used resulting in a reduction in possible codon usage from 64 to 32 and reduces the number of possible stop codons. Position one, therefore, comprises only A and C in the synthetic reaction mixture. For complete randomization of the

second and third positions of the codons the dNTP mixture comprise 25% each of A,G,C and T.

The 3' oligonucleotide randomizing primer was designed such that the last 15 nucleotides of framework 3 and the first 17 nucleotides of framework 4 were kept constant for hybridization. The nucleotides encoding the intervening amino acids, namely amino acids 1-23 of the CDR3 region were randomized using the following primer:
5' (GTTGTCCCTTGGCCCCA n[(A/C)NN]TTTCACACAGTAATA] 3' (Where n=23,antisense strand).

Using a 50% A and 50% C for the first nucleotide position for each anti-codon triplet and 25% each of A, C, G, and T for the second and third nucleotide positions for n=23, complete randomization of the 23 amino acids of the A6 CDR3 is achieved.

2. Synthesis of CDRs comprising 15-23 residues

The primers are adapted by reducing n to 15-23 in the above primer formulae whilst keeping the flanking nucleotides constant.

3. For synthesis of CDR3s comprising 24-33 residues

The primers would be adapted by increasing n to 24-33 in the above primer formula while keeping the flanking nucleotides constant.

Example 17 - Selective Randomization Biasing for 50% Homology to Parental Tyrosine

To achieve approximately 50% homology to wild type at any one position in the A6 dAb CDR3 region during antisense synthesis using the DNA synthesizer, the following example would be used. In the case of tyrosine, which is encoded by TAC or TAT (antisense strand GTA or ATA) the nucleotides would be spiked as follows for the antisense strand.

First anticodon nucleotide position: 80% of A and 20% of C is added to the dNTP solution, and G and T are not added to reduce codon degeneracy.

Second anticodon nucleotide position: 80% T and approximately 6.67% of C, 6.67 of A and 6.67% of G.

Third anticodon nucleotide position the mixture: 80% of A and approximately 6.67% of T and 6.67% of G and 6.7% C.

The calculated probability of tyrosine would thus be 0.8x 0.8 x 0.8 x 100% = 51.2 %. Thus approximately 51% of the chains of the library will contain a wild-type A6 tyrosine in that specified position.

Example 18 - Selective Randomization Biasing for 50% Homology to Parental Serine

Using the same strategy in order to achieve approximately 50% homology to wild type serine at one or more positions, the following example is useful.

Using only A and/or C in the first anticodon position the amino acid serine could have two codons these are AGT, TCT and TCG (antisense ACT, AGA and CGA, respectively). The nucleotide spiking levels would be as follows:

First anticodon nucleotide position: 50% A and 50% C.

Second anticodon nucleotide position: 35.35% C, 35.35% G, 14.65% A and 14.65% T

Third anticodon nucleotide position: 35.35% A, 35.35% T, 14.65% C and 14.65% G.

The probability of producing serine for any given fragment, using this strategy is (1 x [0.3535+0.3535] x [0.3535+0.3535] x 100% = 50 %. Thus, approximately 50% of the chains will have a serine in the selected position.

Example 19 - Selective Randomization Biasing for 50% Homology to Parental Serine

To achieve approximately 10% homology to wild type at any one position in the A6 dAb CDR3 region during antisense synthesis using the DNA synthesizer, the following example can be used. In the case of tyrosine which is encoded by TAC or TAT (antisense strand GTA or ATA) the nucleotides would be spiked as follows for the anti sense strand.

First anticodon nucleotide position: 47% of A and 53% of C is added; G and T are not added to reduce codon degeneracy.

Second anticodon nucleotide position: 47% T and approximately 17.67 % of C, 17.67 of A and 17.67% of G.

Third anticodon nucleotide position: 47% of A and approximately 17.67% of T and 17.67% of G and 17.67% C.

The calculated probability of tyrosine is thus 0.47x 0.47 x 0.47 x 100% = 10.4 %. Thus approximately 10% of the chains of the library will contain a wild-type A6 tyrosine in that specified position.

Example 20 - Selective Randomization Biasing for 50% Homology to Parental Serine

To achieve approximately 90% homology to wild-type amino acids at any positions in the A6 dAb CDR3 region during antisense synthesis using the DNA synthesizer , the following example would be used. In the case of tyrosine which is encoded by TAC or TAT (antisense strand GTA or ATA) the nucleotides would be spiked as follows:

First anticodon nucleotide position: 97% of A and 3% of C is added, G and T are not added to reduce codon degeneracy. For this reason, only A and C are used in the first anticodon position for all 20 naturally occurring amino acids.

Second antcodon nucleotide position: 97% T and approximately 1 % of C, 1% of A and 1% of G.

Third anticodon nucleotide position: 97% of A and approximately 1% of T and 1% of G and 1% C.

The calculated probability of tyrosine would be 0.97x 0.97 x 0.97 x 100% = 91.3 %. Thus approximately 90% of the chains of the library will contain a wild-type A6 tyrosine in that specified position.

Using the approaches in the examples above, approximately 10 % to approximately 90 % of wild type amino acid representation at one or more specified amino acid residues in the A6 CDR3 can be calculated and applied to the DNA synthesizer.

Example 21

Removal of the recombination site

Figure 15 shows a schematic representation of the steps taken to remove the putative recombination site at the 5' end of the A6VH gene. For simplicity only the part of the plasmid spanning from RP (1) primer binding site to FP (2) primer binding site and containing the A6VH gene is shown. 3=Chi.F primer; 4=Chi.R primer (explained below)

The codons for amino acids 3-16 surrounding the recombination site were changed (Figure 15). Briefly, using the Chi.R-5'(CAATTACAAGAAAGTGGTGGCGGACTGGTGCAACCAGGAGGATCCCTGAGACTC)3'/FP and Chi.F-5'(ACTTTCTTGTAATTGGACCTCGGCCTGCGC)3'/RP primers pairs and pSJF-A6VH plasmid as template two 5' and 3' fragments were synthesised by PCR in a total volume of 50 ul containing 10 pmol each of the two primers, 2 mM each of the four dNTPs, 1x buffer and 2.5 units of AmpliTaq™ DNA polymerase (Perkin Elmer). The PCR protocol consisted of an initial denaturation step at 94°C for 3 min followed by 30 cycles of 94°C for 30 sec, 55°C for 30 sec, 72°C for 1 min and a final extension step at 72°C for 10 min. The two fragments were gel purified using the QIAquick Gel Extraction™ kit (QIAGEN), and a larger construct was assembled from the 5' and 3' fragments by performing splice overlap extension (SOE). Briefly, the reaction vial containing both 5' and 3' fragments, 200 μM each of the four dNTPs, 5 μl 10X buffer (NEB), and 2 units of Vent DNA polymerase (NEB) was subjected to 7 cycles of 1 min at 94°C and 2.5 min at 72°C. To

amplify the assembled construct, RP and FP primers were added at a final concentration of 1 pmol/μl and the mixture was subjected to 30 cycles of 1 min at 94°C, 30 s at 55°C, and 1 min at 72°C. The amplified product was purified (QIAquick PCR Purification™ kit) and subsequent sequencing revealed that the desired mutations had been incorporated into the VH.

The present invention may be embodied in other specific forms without departing from the spirit or essential characteristics thereof. Certain adaptations and modifications of the invention will be obvious to those skilled in the art. Therefore, the presently discussed embodiments are considered to be illustrative and not restrictive. It is understood that the claims may refer to aspects or embodiments of the invention that are only inferentially referred to in the disclosure.

Additional References:

Anand, N.N., Dubuc, G., Phipps, J., MacKenzie, C.R., Sadowska, J., Young, N.M., Bundle, D.R. and Narang, S.A. (1991). Synthesis and expression in *Escherichia coli* of cistronic DNA encoding an antibody fragment for a Salmonella serogroup B O-antigen. Gene **100**, 39-44.

Deng, S.-J., MacKenzie, C. R., Hirama, T., Brousseau, R., Lowary, T. L., Young, N. M., Bundle, D. R., and Narang, S. A. 1995. Analysis by surface plasmon resonance of the influence of valence on the ligand-binding affinity and kinetics of an anti-carbohydrate antibody. Proc. Natl. Acad. Sci. U. S. A. **92**, 4992-4996

Jönsson, U., Fägerstam, L., Ivarsson, B., Johnsson, B., Karlsson, R., Lundh, K., Löfås, S., Persson, B., Roos, H., Rönnberg, I., Sjölander, S., Stenberg, E., Ståhlberg, R., Urbaniczky, C., Östlin, H., and Malmqvist, M. (1991). Real-time biospecific interaction analysis using surface plasmon resonance and a sensor chip technology. BioTechniques **11**, 620-627.

Knappik, A. and Pluckthun, A. (1994). An improved affinity tag based on the FLAG peptide for the detection and purification of recombinant antibody fragments. Biotechniques **17**, 754-761.

MacKenzie, C.R., Sharma, V., Brummel, D., Bilous, D., Dubuc, G., Sadowska, J., Young, N. M., Bundle, D. R., and Narang, S. A. 1994. Effect of $C_?$-$C_?$ domain switching on Fab activity and yield in Escherichia coli: synthesis and expression of genes encoding two anti-carbohydrate Fabs. Bio/Technology **12**: 390-395.

MacKenzie, C.R., Hirama, T., Deng, S.-J., Bundle, D.R., Narang, S.A., and Young. N.M. (1996). Analysis by surface plasmon resonance of the influence of valence on the ligand-binding affinity and kinetics of an anti-carbohydrate antibody. J. Biol. Chem. **271**, 1527-1533

Miceli, R.M., Degraff, M.E. and Fischer, H.D. (1994).J. Immunol. Methods **167**, 279-287.

Harrison, J.L, Williams, S.C., Winter, G. and Nissim, A. (1996). Screening of phage antibody libraries. Methods in Enzymology **267**, 83-109.

McAfferty, J., Griffiths, A.D., Winter, G. and Chiswell, D.J. (1990). Nature **348**, 552-554.

Messing, J. (1983). New M13 vectors for cloning. Methods in Enzymology 101, 20-???.

Narang, S.A., Yao, F.-L., Michniewicz, J.J., Dubuc, G., Phipps, J. and Somorjai, R.L. (1987). Hierarchical strategy for protein folding and design: synthesis and expression of T4 lysozyme gene and two putative folding mutants. Protein Engineering **1**, 481-485.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989). Molecular Cloning: A Laboratory Manual (2^{nd} edition). Cold Spring Laboratory Laboratory, Cold Spring Harbor, NY.

Sanger, F., Nicklen, S. and Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA, **74**, 5463.

Tung, W.L. and Chow, K.-C. (1995). A modified medium for efficient transformation of *E.coli*. Trends in Genetics **11**, 128-129

Vierra, O. and Messing, J. (1982). The pUC plasmids , an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene **19**, 259-268.

Zacher, A.N., Stock, C.A., Golden, J.W. and Smith, G.P. (1980). Gene **9**, 127-140.

Anand, N. N., Dubuc, G., Phipps, J., MacKenzie, C. R., Sadowska, J., Young, N. M., Bundle, D. R., and Narang, S. A. (1991). Synthesis and expression in Escherichia coli of cistronic DNA encoding an antibody fragment specific for a Salmonella serotype B O-antigen. *Gene* **100**:39-44.

Arbabi, G. M., Desmyter, A., Wyns, L., Hamers, R., and Muyldermans, S. (1997). Selection and identification of single domain antibody fragments from camel heavy-chain antibodies. *FEBS Lett* **414**:521-526.

Bodenhausen, G. and Ruben, D. J.; Chem. Phys. Lett., 69, 185-188, 1980.

Cai, X. and Garen, A. (1996). A melanoma-specific VH antibody cloned from a fusion phage library of a vaccinated melanoma patient. *Proc Natl Acad Sci U S A* **93**:6280-6285.

Chothia, C. and Lesk, A. M. (1987). Canonical structures for the hypervariable regions of immunoglobulins. *J Mol Biol* **196**:901-917.

Chothia, C., Novotny, J., Bruccoleri, R., and Karplus, M. (1985). Domain association in immunoglobulin molecules. The packing of variable domains. *J Mol Biol* **186**:651-663.

Clackson, T., Hoogenboom, H. R., Griffiths, A. D., and Winter, G. (1991). Making antibody fragments using phage display libraries. *Nature* **352**:624-628.

Constantine, K. L., Goldfarb, V., Wittekind, M., Anthony, J., Ng, S. C., and Mueller, L. (1992). Sequential 1H and 15N NMR assignments and secondary structure of a recombinant anti-digoxin antibody VL domain. *Biochemistry* **31**:5033-5043.

Constantine, K. L., Goldfarb, V., Wittekind, M., Friedrichs, M. S., Anthony, J., Ng, S. C., and Mueller, L. (1993). Aliphatic 1H and 13C resonance assignments for the 26-10 antibody VL domain derived from heteronuclear multidimensional NMR spectroscopy. *J Biomol NMR* **3**:41-54.

Davies, J. and Riechmann, L. (1994). 'Camelising' human antibody fragments: NMR studies on VH domains. *FEBS Lett* **339**:285-290.

Davies, J. and Riechmann, L. (1995). Antibody VH domains as small recognition units. *Biotechnology N Y* **13**:475-479.

Davies, J. and Riechmann, L. (1996b). Affinity improvement of single antibody VH domains: residues in all three hypervariable regions affect antigen binding. *Immunotechnology* **2**:169-179.

Davies, J. and Riechmann, L. (1996a). Single antibody domains as small recognition units: design and in vitro antigen selection of camelized, human VH domains with improved protein stability. *Protein Eng* **9**:531-537.

Decanniere, K., Desmyter, A., Lauwereys, M., Ghahroudi, M. A., Muyldermans, S., and Wyns, L. (1999). A single-domain antibody fragment in complex with RNase A: non-canonical loop structures and nanomolar affinity using two CDR loops. *Structure* **7**:361-370.

Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J. and Bax, A., J. Biomol. NMR 6, 277-293, 1995.

Deng, S. J., MacKenzie, C. R., Hirama, T., Brousseau, R., Lowary, T. L., Young, N. M., Bundle, D. R., and Narang, S. A. (1995). Basis for selection of improved carbohydrate-binding single-chain antibodies from synthetic gene libraries. *Proc Natl Acad Sci U S A* **92** :4992-4996.

Deng, S. J., MacKenzie, C. R., Sadowska, J., Michniewicz, J., Young, N. M., Bundle, D. R., and Narang, S. A. (1994). Selection of antibody single-chain variable fragments with improved carbohydrate binding by phage display. *J Biol Chem* **269**:9533-9538.

Desmyter, A., Transue, T. R., Ghahroudi, M. A., Thi, M. H., Poortmans, F., Hamers, R., Muyldermans, S., and Wyns, L. (1996). Crystal structure of a camel single-domain VH antibody fragment in complex with lysozyme [see comments]. *Nat Struct Biol* **3**:803-811.

Hamers, C. C., Atarhouch, T., Muyldermans, S., Robinson, G., Hamers, C., Songa, E. B., Bendahman, N., and Hamers, R. (1993). Naturally occurring antibodies devoid of light chains. *Nature* **363**:446-448.

Harrison, J. L., Williams, S. C., Winter, G., and Nissim, A. (1996). Screening of phage antibody libraries. *Methods Enzymol* **267**:83-109.

Johnson, B. A. and Blevins, R. A., J. Chem., Phys., 29, 1012-1014, 1994.

Knappik, A. and Pluckthun, A. (1994). An improved affinity tag based on the FLAG peptide for the detection and purification of recombinant antibody fragments. *BioTechniques* **17**:754-761.

Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**:680-685.

Lauwereys, M., Arbabi, G. M., Desmyter, A., Kinne, J., Holzer, W., De Genst, E., Wyns, L., and Muyldermans, S. (1998). Potent enzyme inhibitors derived from dromedary heavy-chain antibodies. *EMBO J* **17**:3512-3520.

Lowman, H. B., Bass, S. H., Simpson, N., and Wells, J. A. (1991). Selecting high-affinity binding proteins by monovalent phage display. *Biochemistry* **30**:10832-10838.

MacKenzie, R. and To, R. (1998). The role of valency in the selection of anti-carbohydrate single-chain Fvs from phage display libraries. *J. Immunol. Methods* **220**, 39-49.

MacKenzie, C. R., Sharma, V., Brummell, D., Bilous, D., Dubuc, G., Sadowska, J., Young, N. M., Bundle, D. R., and Narang, S. A. (1994). Effect of C lambda-C kappa domain switching on Fab activity and yield in Escherichia coli: synthesis and expression of genes encoding two anti-carbohydrate Fabs. *Biotechnology N Y* **12**:390-395.

Miceli, R. M., Degraaf, M. E., and Fischer, H. D. (1994). Two-stage selection of sequences from a random phage display library delineates both core residues and permitted structural range within an epitope. *J Immunol Methods* **167**:279-287.

Morea, V., Tramontano, A., Rustici, M., Chothia, C., and Lesk, A. M. (1998). Conformations of the third hypervariable region in the VH domain of immunoglobulins. *J Mol Biol* **275**:269-294.

Muyldermans, S., Atarhouch, T., Saldanha, J., Barbosa, J. A., and Hamers, R. (1994). Sequence and structure of VH domain from naturally occurring camel heavy chain immunoglobulins lacking light chains. *Protein Eng* **7**:1129-1135.

Narang, S. A., Yao, F. L., MichnieWics, J. J., Dubuc, G., Phipps, J., and Somorjai, R. L. (1987) Protein Eng. **1**, 481-485.

Narang, S. A., Brummell, D. A., Sharma, V., Dubuc, G., MacKenzie, C. R., MichnieWics, J. J., Sadowska, J., Anand, N., Young, N. M., and Bundle, D.R. (1991) Nucleic Acids Res. Symp. Ser. **24**, 173-179.

Nissim, A., Hoogenboom, H. R., Tomlinson, I. M., Flynn, G., Midgley, C., Lane, D., and Winter, G. (1994). Antibody fragments from a 'single pot' phage display library as immunochemical reagents. *EMBO J* **13**:692-698.

Padlan, E. A. (1994). Anatomy of the antibody molecule. *Mol Immunol* **31**:169-217.

Piotto, M., Saudek, V., and Sklenar, V. (1992) *J. Biomol. NMR 2,* 661-665.

Sklenar, V., Piotto, M., Leppik, R., and Saudek, V. (1993) *J. Magn. Reson. A102,* 241-245.

Reiter, Y., Schuck, P., Boyd, L. F., and Plaksin, D. (1999). An antibody single-domain phage display library of a native heavy chain variable region: isolation of functional single-domain VH molecules with a unique interface. *J Mol Biol* **290**:685-698.

Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* **74**:5463-5467.

Sheriff, S. and Constantine, K. L. (1996). Redefining the minimal antigen-binding fragment [news; comment]. *Nat Struct Biol* **3**:733-736.

Slatter, M., Schleucher, J. and Griesinger, C., Progr. NMR Spectrosc., 1999, 34, 93-158.

Spinelli, S., Frenken, L., Bourgeois, D., de Ron, L., Bos, W., Verrips, T., Anguille, C., Cambillau, C., and Tegoni, M. (1996). The crystal structure of a llama heavy chain variable domain [letter] [see comments]. *Nat Struct Biol* **3**:752-757.

Tung, W. L. and Chow, K. C. (1995). A modified medium for efficient electrotransformation of E. coli. *Trends Genet* **11**:128-129.

Vu, K. B., Ghahroudi, M. A., Wyns, L., and Muyldermans, S. (1997). Comparison of llama VH sequences from conventional and heavy chain antibodies. *Mol Immunol* **34**:1121-1131.

Ward, E. S., Gussow, D., Griffiths, A. D., Jones, P. T., and Winter, G. (1989). Binding activities of a repertoire of single immunoglobulin variable domains secreted from Escherichia coli [see comments]. *Nature* **341**:544-546.

Wishart, D. S. and Sykes, B. D., J Biomol NMR, 4, 171-180, 1994.

Wishart, D. S., Bigam, C. G., Yao, J., Abildgaard, F., Dyson, H. J., Oldfield, E., Markley, J. L. and Sykes, B. D., J. Biomol. NMR, 6 135-140, 1995.

Davies, J., and Riechmann, L. 1996. Affinity improvement of single antibody VH domains: residues in all three hypervariable regions affect antigen binding. Immunotechnology **2**, 169-179.

Davies, J. and Riechmann, L. (1996). Single antibody domains as small recognition units: design and *in vitro* antigen selection of camelized, human VH domains with improved protein stability. *Protein Eng.* **9**, 531-537.

Sambrook, J. Fritsch E. F. and Maniatis T (1989). "Molecular Cloning: A laboratory Manual (2^{nd} ed.)," Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Dan, M., Earley, E. M., Griffin, M. C., Maiti, P. K., Prashar, A. K., Yuan, X. Y., Friesen, A. D., and Kaplan, H. A. 1995. Human monoclonal antibody BT32/A6 and a cell cycle-independent glioma-associated surface antigen. J. Neurosurg. **82**, 475-480.

Riechmann, L. (1996). Rearrangement of the former VL interface in the solution structure of a camelised, single antibody VH domain. *J Mol Biol* **259**:957-969.

Riechmann, L. and Davies, J. (1995). Backbone assignment, secondary structure and protein A binding of an isolated, human antibody VH domain. *J Biomol NMR* **6**:141-152.

Messing, J. (1983). New M13 vectors for cloning. Methods in Enzymology 101, 20.

Vierra, O. and Messing, J. (1982). The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene **19**, 259-268.

Zacher, A. N., Stock, C. A., Golden, J. W., and Smith, G. P. (1980). Gene **9**, 127-140.

McAfferty, J., Griffiths, A. D., Winter, G. and Chiswell, D. J. (1990). Nature, **348**,

Company Names

Amersham Pharmacia Biotech, Baie d'Urfé, QC, Canada
NUNC, Roskilde, Denmark
Bio-Rad Laboratories, Missisauga, ON
PE Biosystems, Mississauga, ON, Canada